وزارة التعليم العالي والبحث العلمي

جامعة الشهيد الشيخ العربي تبسي – تبسة-

كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير

قسم العلوم الاقتصادية

الرقم التسلسلي: .....

مذكرة مقدمة لاستكمال متطلبات نيل شهادة ماستر أكاديمي (ل م د)

الميدان: علوم اقتصادية، تسيير وعلوم تجارية

الشعبة: علوم اقتصادية

ماستر تخصص: اقتصاد كمي

المذكرة موسومة بـ:

التنبؤ بالسلوك الإستهلاكي للزبون بإستخدام طرق تعلم الآلة لعينة من العلامات التجارية

من إعداد الطالب(ة):

براهمي ريم

تحت اشراف:

أ_ حميدان ربيع

أعضاء لجنة المناقشة

الاسم واللقب	الرتبة العلمية	الصفة
د. عمير حمه	أستاذ محاضر أ	رئيسا
أ_ حميدان ربيع	أستاذ مساعد أ	مشرفا ومقررا
د. بوعمرة حسن	أستاذ محاضر ب	عضوا مناقشا

السنة الجامعية 2024-2025

ملخص الدراسة:

يعتبر السلوك الإستهلاكي للزبون عنصرا مهما لفهم العلاقة بين الزبون والشركات وتوطيدها أكثر، ومن بين السلوكيات التي تشكل خطرا وتهديدا بالنسبة لهذه الشركات هو سلوك المغادرة، إذ يساعد التنبؤ به على كشفه مسبقا مما يسمح لها من الحد من مغادرة الزبائن و زيادة لائهم وإتخاذ قرارات مسبقة مبنية على أسس علمية ونتائج دقيقة و حماية نفسها من زيادة تكاليفها وخسارة زبائنها.

تسعى هذه الدراسة إلى بناء نموذج قادر على التنبؤ بالسلوك الإستهلاكي للزبائن من خلال تصنيفهم إما مغادرتهم أو بقائهم بإستخدام تقنيات تعلم الآلة، حيث تم تطبيق المنهج الوصفي التحليلي إذ تم الإعتماد في هذه الدراسة على قاعدة بيانات حقيقية لأحد شركات الإتصالات تم إحضارها من موقع Kaggle وتم تجريب 4 نماذج لتعلم الآلة وهي(شجرة القرارDecision Tree، التدرج المعززGradient boosting، الإنحدار اللوجستي Logistic Regression، الغابات العشوائية Random Forest) والمقارنة بينهم من خلال المؤشرات العامة (Recall، Precision، Accuracy، F1-Scores) وذلك بهدف إختيار النموذج الأكثر كفاءة في التنبؤ، وتوصلت هذه الدراسة إلى أن أفضل نموذج من حيث نسبة الدقة هو نموذج التدرج المعزز والذي أعطى نسبة دقة 83% إضافة إلى أن نتيجة الإنحراف المعياري كانت 0.01 مما يشير إلى أن البيانات كانت جد مستقرة .

الكلمات المفتاحية: السلوك الإستهلاكي، الزبائن، تعلم الآلة، التنبؤ

Abstract

Study Summary: Consumer behavior is considered an important element in understanding and strengthening the relationship between customers and companies. Among the behaviors that pose a risk and threat to these companies is the behavior of leaving, as predicting it helps in uncovering it in advance, allowing companies to reduce customer departures, increase their loyalty, make proactive decisions based on scientific foundations and accurate results, and protect themselves from increased costs and loss of customers. This study seeks to build a model capable of predicting consumer behavior by classifying customers as either leaving or staying using machine learning techniques. The descriptive analytical approach was applied, relying on a real database from one of the telecommunications companies obtained from Kaggle, where four machine learning models were tested (Decision Tree, Gradient Boosting, Logistic Regression, Random Forest) and compared using general indicators (Recall, Precision, Accuracy, F1-Scores) with the aim of selecting the most efficient model for prediction،This study concluded that the best model in terms of accuracy is the boosted gradient model, which achieved an accuracy rate of 83%. Additionally, the standard deviation was 0.01, indicating that the data was quite stable.

Keywords: consumer behavior, customers, machine learning, prediction.

شكر و عرفان

الحمد لله حمدا كثيرا حتى يبلغ الحمد منتهاه والصلاة والسلام على أشرف مخلوق أناره الله وإصطفاه سيدنا محمد عليه أفضل الصلاة وأتم التسليم أما بعد فمن لم يشكر الناس لم يشكر الله

لا يسعني إلا أن أتقدم بالشكر والتقدير وعظيم الإمتنان لأستاذي الفاضل حميدان ربيع والذي تفضل بإشرافه على هذا العمل وكان سندا في كل مراحل إنجازه وعلى إرشاده وتوجيهاته القيمة ولما قدمه من مساعدة لإتمام هذا العمل المتواضع

وأشكر أيضا الأستاذ عطية عز الدين وذلك لما قدمه من إفادات ونصائح قيمة خاصة في الجانب المنهجي لهذه الدراسة

ولا يفوتني أن أتوجه بجزيل الشكر والإمتنان لجميع الأساتذة الكرام الذين درسوني طوال مشواري في الجامعة فقد كانت بصماتهم واضحة لوصولي إلى هذه المرحلة شكرا لكم جميعا على ما قدمتموه، ولكم مني فائق الإحترام

الحمد لله الذي أجرى سنوات دراستي حتى توالت وأكرمني بنعمة التعلم والتوجيه من الأساتذة أسأل الله أن يبارك في علمهم ويجزيهم خير الجزاء

وختاما آمل أن يكون هذا العمل صدقة جارية ينتفع بها كل من أراد البحث أو الإطلاع على موضوع المذكرة وأن يكتب الله له القبول والنفع .

إهداء

((وآخر دعواهم أن الحمد لله رب العالمين))

الحمد لله الذي ما تم جهد ولا ختم سعي إلا بفضله وما سلكنا البدايات إلا بتيسيره وما بلغنا النهايات إلا بفضله فالحمد لله حبا وشكرا وإمتنانا والحمد لله على البدء والختام

لم تكن الرحلة قصيرة ولا ينبغي لها أن تكون كذلك ، لم يكن الحلم قريبا بل كان الطريق إلى تحقيقه محفوفا بالعثرات لكني فعلتها بتوفيق من الله أولا وإصرارا مني ودعما من عائلتي ثانيا

أهدي تخرجي وثمرة جهدي وحصاد ما زرعته خلال 17 سنة من الدراسة بداية من السنة الأولى إبتدائي 2008 إلى غاية 2025 السنة ثانية ماستر سنة التخرج إلى أول شخص أخذني إلى المدرسة في أول يوم إلى من أفنى عمره لأجلي وإخوتي وإنتظر هذه اللحظة كما أنتظرتها إلى من غادر الدنيا بجسده لكنه مازال حيا في قلبي إلى أول من علمني حروف الحياة، من أفتقده في مواجهة الصعاب ولم تمهله الدنيا لأرتوي من حنانه يا من تمنيت أن تكون بين الحاضرين فكنت الأقرب إلى روحي رغم الغياب إلى فقيد قلبي أبي العزيز رحمه الله

إلى اليد الخفية والحنونة التي أزالت عني طريق الأشواك وأشعرتني بالأمان وغرست فيا حب العلم والتعلم والتي كانت دعواتها سر نجاحي وسندي الأول التي دعمتني لإكمال مشوار الدراسة والتي ستبقى كلماتها ونصائحها نجوما أهتدي بها اليوم والغد وإلى الأبد أمي الغالية حفظها الله

إلى من شد الله عضدي بهم و وهبني نعمة وجودهم إلى العقد المتين وضلعي الثابت إلى تلك النجوم التي دائما ما كانت تنير طريقي بنصائحهم وحبهم إلى من تحمل المسؤولية بعد رحيل أبي وكان نعم الأخ والأب في نفس الوقت إخوتي وأخواتي حفظهم الله وجعل حياتهم مليئة باليمن والبركة

إلى نفسي الطموحة التي صمدت لأجل النجاح وكانت سحابا ممطرا في السنوات العجاف شكرا لكي لأنك لم تخذلي ذاتكي و واصلتي إضاءة نفسكي بعد الإنطفاء في كل مرة

إلى من عرفني بهم القدر وتميزوا بالإخاء ولم تحلو المسيرة الجامعية إلا بوجودهم وكانوا رفقاء للدرب خلال السنوات الخمس وكانوا الضحكة التي خففت عناء الطريق ستبقى ذكرياتنا محفورة في القلب مهما باعدت بنا الأيام دفعة إقتصاد كمي 2025 شكرا لكم على تلك الأيام الحلوة و وفقكم الله وسدد خطاكم.

شكر وعرفان .....
إهداء .....
قائمة الجداول.....ب
قائمة الصور.....ج
مقدمة عامة 1
الفصل الأول : الإطار النظري والدراسات السابقة 4
المبحث الأول: مدخل إلى السلوك الإستهلاكي للزبون 4
المطلب الأول: مفهوم السلوك الإستهلاكي للزبون.....4
المطلب الثاني: محددات السلوك الإستهلاكي.....8
المطلب الثالث: سلوكيات الزبائن الممكن التنبؤ بها 13
المبحث الثاني: طرق التنبؤ بإستخدام التعلم الآلي.....16
المطلب الأول: عموميات حول التنبؤ.....17
المطلب الثاني: مدخل إلى تعلم الألة 19
المطلب *الثالث : طرق التنبؤ بإستخدام تعلم الآلة* 25
المبحث الثالث: الدراسات السابقة.....27
المطلب الأول: الدراسات العربية.....28
المطلب الثاني: الدراسات الأجنبية.....30
الفصل الثاني : الفصل التطبيقي 32
المبحث الأول: إستكشاف بيانات الدراسة.....32
المطلب الأول: التعريف بمتغيرات الدراسة ومصدر جمع البيانات 32
المطلب الثاني: الأدوات وبيئة التطوير المستخدمة في الدراسة.....35
المطلب الثالث: خطوات بناء النموذج وأدوات تقييمه.....37
المبحث الثاني : تحليل النتائج وتفسيرها 43
المطلب الأول: إستكشاف البيانات و معالجتها 44
المطلب الثاني : بناء و تقييم النموذج.....51
المطلب الثالث: المقارنة بين الدراسة الحالية والدراسة السابقة 66
خاتمة.....68
قائمة المراجع.....70

الفصل الثاني:

جدول1.2: وصف متغيرات الدراسة ص33

جدول2.2: جدول يوضح مصفوفة الإرتباك أو الإلتباس Confusion Matrix.....ص40

جدول3.2: عرض أول5 مشاهدات من قاعدةالبيانات.....ص44

جدول4.2: إحصائيات البيانات قبل المعالجة.....ص46

جدول 5.2: إحصائيات البيانات بعد المعالجة.....ص47

جدول 6.2: نتائج مصفوفة التحقق المتقاطع حسب نموذج شجرة القرار.....ص51

جدول 7.2: مؤشرات الأداء العامة حسب نموذج شجرة القرار.....ص52

جدول8.2: نتائج مصفوفة التحقق المتقاطع حسب نموذج شجرة القرار.....ص53

جدول 9.2: نسبة مساهمة المتغيرات في بناء نموذج شجرة القرار.....ص54

جدول 10.2: نتائج مصفوفة اللاإلتباس حسب نموذج الإنحدار اللوجستي.....ص55

جدول 11.2: نتائج المؤشرات العامة لنموذج الإنحدار اللوجستي.....ص56

جدول12.2: نتائج مصفوفة التحقق المتقاطع حسب نموذج الإنحدار اللوجستي.....ص57

جدول13.2: نسبة مساهمة المتغيرات في بناء نموذج الإنحدار اللوجستي.....ص68

جدول14.2: نتائج مصفوفة الإلتباس حسب نموذج التدرج المعزز.....ص69

جدول 15.2: نتائج المؤشرات العامة لنموذج التدرج المعزز.....ص59

جدول16.2: نتائج مصفوفة التحقق المتقاطع لنموذج التدرج المعزز.....ص60

جدول17.2: نسبة مساهمة المتغيرات في بناء نموذج التدرج المعزز.....ص61

جدول18.2 : نتائج مصفوفة الإلتباس لنموذج الغابات العشوائية.....ص62

جدول 19.2: المؤشرات العامة لنموذج الغابات العشوائية.....ص63

جدول 20.2: نتائج مصفوفة التحقق المتقاطع لنموذج الغابات العشوائية.....ص64

جدول 21.2: نسبة مساهمة المتغيرات في بناء نموذج الغابات العشوائية.....ص64

جدول 22.2: مؤشرات المقارنة بين الدراستين.....ص65

الفصل الاول:

الشكل 1.1: مصفوفة زبائن المنظمة.....ص6

الشكل 2.1: هرم موسلو للحاجات.....ص10

الشكل 0.1: منحنى المنفعة الحدية والمنفعة الكلية.....ص12

الشكل4.1: مراحل إتخاذ القرار الشرائي.....ص14

الشكل5.1: مراحل التنبؤ.....ص18

الشكل 6.1: توضيح لطريقة التجميع.....ص23

الشكل 7.1: رسم توضيحي لطريقة التعلم المعزز.....ص*24*

الفصل الثاني:

الشكل 8.2: خطوات بناء النموذج.....ص38

الشكل9.2: كيفية عمل مصفوفة الإلتباس cross validation.....ص42

الشكل10.2: إستيراد المكتبات.....ص43

الشكل11.2: إستيراد قاعدة البيانات ص43

الشكل 12.2: مصفوفة الإرتباطات.....ص*48*

الشكل13.2: المتغيرات التي تم إختيارها بالإعتماد على مصفوفة الإرتباطات.....ص49

الشكل14.2: إستيراد نموذج شجرة القرار.....ص50

الشكل 15.2: الأمر الخاص بإستخراج مصفوفة confusion matrix.....ص51

الشكل16.2: أمر مصفوفة الإلتباس الخاصة بنموذج الإنحدار اللوجستي.....ص55

الشكل17.2: أمر إستيراد نموذج التدرج المعزز.....ص58

الشكل18.2:أمر إستخراج مصفوفة الإلتباس لنموذج التدرج المعزز.....ص59

الشكل19.2: إستيراد نموذج الغابات العشوائية.....ص62

الشكل20.2: أمر إستخراج مصفوفة الإلتباس لنموذج الغابات العشوائية.....ص62

مقدمة عامة

نظرا إلى ما آل إليه العالم من تحولات جذرية في الإقتصاد والتكنولوجيا، فإن النجاح لم يعد حكرا على الشركات التي تقدم منتجات أو خدمات ذات جودة عالية فقط بل أصبح مرتبطا بصفة مباشرة بمدى قدرتها على فهم إحتياجات ورغبات عملائها بعمق، وأصبحت كل الشركات تتنافس لمعرفة ثغرات السوق بغية تحسين منتجاتها وجعلها أكثر تميزا وهذا ما يجعل الزبائن أمام تنوع وتعدد في الخيارات، خاصة في عالم تسوده المنافسة والتغير المستمر في أوضاع السوق، لهذا أصبحت الشركات والمؤسسات تسعى إلى فهم السلوك الإستهلاكي للزبائن بإعتباره مفتاحا أساسيا لنجاحهم فالزبون ليس مجرد فرد يقوم بإتخاذ قرار الشراء بشكل عشوائي وإنما هو فرد له تصرف وسلوك معقد ومتعدد الأبعاد و تؤثر فيه عدة عوامل و تحركه مجموعة من الدوافع سواء كانت دوافع نفسية، إجتماعية، ثقافية، أو غيرها لتساهم في إتخاذ قراراته اليومية فدراسة سلوك هذا الفرد لا تقتصر على معرفة ما يشتريه الأفراد فقط بل يصل ذلك إلى فهم كيف يتم إتخاذهم لقرار الشراء ومعرفة العوامل التي تلعب دورا في تحديد تفضيلاتهم وكذلك طريقة تفاعلهم مع المنتجات والخدمات؛

حيث يعتبر فهم سلوكهم حجر الأساس الذي تعتمد عليه الشركات بدءا من تصميم المنتجات وتحديد الأسعار وتطوير إستراتيجياتها وصولا إلى إختيار قنوات التوزيع و وسائل الترويج الأكثر فعالية وذلك من خلال تحليل الأنماط الشرائية والتوجهات الإستهلاكية و التي تساعد الشركات على التنبؤ بإحتياجات السوق والإستجابة بسلاسة للتغيرات من خلال تحليل البيانات المتعلقة بسلوك الزبائن مثل سجلات الشراء أو التفاعل على مواقع التواصل الإجتماعي بإستعمال تقنيات تعلم الآلة وبناء نماذج تنبؤية تمنحها ميزة تنافسية تزيد من فرص نجاحها.

1. الإشكالية الرئيسية: كيف يمكن التنبؤ بالسلوك الإستهلاكي للزبائن لشركة إتصالات بمقاطعة كاليفورنيا بإستخدام تعلم الآلة؟

2. الأسئلة الفرعية:

- ماهو مفهوم السلوك الإستهلاكي للزبون وماهي أهم تقنيات تعلم الآلة المستخدمة في التنبؤ؟
- كيف يمكن إستخدام خوارزميات التصنيف للتنبؤ بسلوك الزبون؟
- هل يمكن لأداء النموذجين Random Foresو GBM التفوق على أداء نموذج Decision Tree و نموذج Logistic Regression في مهمة تصنيف سلوك الزبون الاستهلاكي؟

3. الفرضية الرئيسية:

تقنيات تعلم الآلة تستطيع التنبؤ بالسلوك الإستهلاكي للزبائن بدقة بالإعتماد على بياناتهم في شركة الإتصالات.

4. الفرضيات الفرعية:

–أداء النموذجين Random Forest و GBM يفوق أداء نموذج Decision Tree و نموذج Logistic Regression في مهمة تصنيف سلوك الزبون الإستهلاكي.

- إستخدام تقنيات Cross Validation يُحسن من مصداقية نتائج التقييم ويقلل من التحيز في اختيار النموذج الأنسب.

-نموذج GBM يحقق أعلى اداء مقارنة بباقي النماذج المستخدمة في مهمة تصنيف الزبائن حسب سلوكهم الاستهلاكي من حيث مؤشرات الدقة .(Accuracy, F1-score, Precision et Recall).

5. أهمية الدراسة:

- تساعد هذه الدراسة على تحسين من عملية إتخاذ القرار داخل المؤسسات والمساهمة في تحسين إستراتيجيات التسويق إستنادا إلى نتائج نموذج التنبؤ بسلوك الزبون الإستهلاكي.

- التنبؤ بسلوك الزبائن الإستهلاكي يساهم في تطوير وتحسين تجربتهم ويزيد من ولائهم لعلامة أو شركة معينة.

- التنبؤ بالسلوك الإستهلاكي للزبون يساعد المؤسسات الإقتصادية على إتخاذ قرارات ناجعة كما يقلل من تكاليفها ويحقق لها التنمية المستدامة.

- قدرة تقنيات تعلم الآلة في التعامل مع كم هائل من البيانات ومعالجتها وإكتشاف الثغرات في البيانات مثل العلاقات غير الواضحة بين المتغيرات.

6. أهداف الدراسة:

- دراسة المفاهيم الأساسية التي تساعد على فهم السلوك الإستهلاكي للزبون؛

- التعرف على أهم خوارزميات تعلم الآلة التي يمكن إستخدامها في بناء نموذج تنبؤ بسلوك الزبائن الإستهلاكي؛

- بناء نموذج تنبؤ بسلوك الزبائن الإستهلاكي بناءا على بيانات خاصة بشركة إتصالات كاليفورنيا؛

- تقييم أداء النموذج والتأكد من أنه يمكن الإعتماد عليه في التنبؤ من خلال معايير خاصة بقياس الدقة مثل معيار Accuracy؛

-مقارنة بين عدة نماذج لتعلم الآلة وإختيار الأفضل بينها من حيث الدقة.

7. منهج اللدراسة:

تم الإعتماد في هذه الدراسة على منهجين مختلفين، المنهج الأول هو المنهج الوصفي وذلك بهدف وصف البيانات والتعرف على أهم المفاهيم والمصطلحات المتعلقة بالدراسة أما المنهج الثاني فهو المنهج التحليلي الكمي بهدف تحليل بيانات الزبائن وسماتهم بإستخدام تقنيات تعلم الآلة والتنبؤ بسلوكهم في المستقبل.

الفصل الأول : الإطار النظري والدراسات السابقة

لقد أصبح فهم سلوك الزبائن أمرا بالغ الأهمية في وقتنا الراهن وذلك لما يعود به من نجاح وإستمرارية على المؤسسات الإقتصادية وحتى تستمر العلاقة بين الزبائن والمؤسسات الإقتصادية يجب عليها هذه الأخيرة دراسة سلوك الزبائن و معرفة أهم المتغيرات التي تؤثر فيه والعمل على كسب الزبون والحفاظ عليه من خلال دراسة هذه المتغيرات التي ستساعد على التنبؤ بسلوك الزبائن، ومع توفر أدوات تعلم الآلة أصبحت المؤسسات تملك القدرة على التنبؤ بسلوك الزبائن عن طريق تحليل بياناتهم، في هذا الفصل سيتم التعرف على مفهوم السلوك الإستهلاكي للزبون وأهم العوامل والمحددات التي تؤثر على سلوكه إضافة إلى مفهوم التنبؤ ومفهوم تعلم الآلة ذكر أهم الخوارزميات التي يمكن إستخدامها للتنبؤ بالسلوك الإستهلاكي للزبون.

المبحث الأول: مدخل إلى السلوك الإستهلاكي للزبون

تعتبر دراسة السلوك الإستهلاكي للزبون أساسا لفهم طريقة تفاعلهم مع المنتجات أو الخدمات وتعاملهم معها وكيفية إتخاذ قرارتهم بناء على أسباب متعددة، يمكن أن تكون إجتماعية أو إقتصادية ويعتبر فهم هذه الأسباب أو المحددات وتحليلها أمرا هاما لفهم سلوك الزبائن.

المطلب الأول: مفهوم السلوك الإستهلاكي للزبون

السلوك عامة يعني الأفعال أو ردود الأفعال التي يصدرها الإنسان ويعبر بها عن قبوله أو رفضه لأمر معين، ويختلف هذا السلوك من فرد لآخر وفقا لعدة عوامل سواء نفسية، إجتماعية، إقتصادية، ثقافية،.. إلخ وبما أن الزبون يعتبرا عنصرا جوهريا في السوق فإن جميع الشركات تسعى لفهم وتحليل سلوكه بهدف الحفاظ على ولائه وزيادة مبيعاتها وقبل التطرق إلى تعريف السلوك الإستهلاكي للزبون وجب تعريف الزبون أولا ثم تعريف السلوك الإستهلاكي للزبون حيث يمكن تعريفهما كما يلي على أنه:

أولا: تعريف الزبونClient

هو الفرد أو الكيان الذي يقوم بعملية الشراء من طرف مؤسسة تجارية أوشركة ويمكن أن يكون الزبون عميلا فرديا يشتري بهدف إشباع رغبات شخصية أو عميلا تجاريا (الغامدي، 2023)؛

كما يعرف الزبون أيضا بأنه المشتري للبضاعة من أي سوق أو مركز تجاري أي هو الفرد الذي يقبل على شراء السلعة المعروضة في السوق (arkam، 2017)، و يمكن تعريفه أيضا على أنه" ذلك الشخص الخارجي الذي يتقدم لشراء منتجات المؤسسة أو الحصول على خدماتها" (إيمان، 2019_2020، صفحة

26)، وعرفه يوسف الطائي وهاشم فوزي دباس العبادي بأنه ذلك الشخص الذي يقوم بإقتناء منتجات المؤسسة سواء من السوق الصناعي أو من داخل المؤسسة بهدف تلبية رغباته (الطائي و هشام فوزي العبادي، 2009) ويمكن القول أن الزبون هو الشخص الذي يقوم بتنفيذ عملية الشراء سواء لمصلحته أي من أجل إستهلاكه الشخصي أو لمصلحة غيره.

ثانيا: تعريف السلوك الإستهلاكي

إختلفت آراء الباحثين حول إعطائه تعريفا محددا إلا أنه يمكن تعريفه بأنه مجموعة من التصرفات والأفعال التي يقوم بها الزبون عند القيام بشراء منتج معين أو أنه الطريقة أو النمط الذي يتبعه الفرد للبحث عن السلع أو الخدمات التي يرى أنها تلبي إحتياجاته وتشبع رغباته ومن ثم ينفذ عملية الشراء؛ ويعرف السلوك الإستهلاكي بأنه: " عبارة عن نشاطات يؤديها الأفراد أثناء إختيارهم وشرائهم لسلع وخدمات، وذلك من أجل إشباع رغبات وسد حاجات مطلوبة لديهم " (الجريسي، 1427هـ، صفحة 44)، كما يمكن أيضا تعريف السلوك الإستهلاكي على أنه "ذلك السلوك الذي يبرزه المستهلك في البحث عن وشراء أو إستخدام السلع أو الخدمات أو الأفكار أو الخبرات التي يتوقع أنها ستشبع رغباته أو حاجاته حسب الإمكانات الشرائية المتاحة " (معراوي، 2020، صفحة 2)

ثالثا: تصنيف سلوك الزبائن

يمكن تصنيف سلوك الزبائن إلى 4 أنماط يمكن للزبون أن يتعامل بها وهي:

أ- على أساس الخصائص الشخصية:

- الزبون العاطفي: يتعامل هذا الزبون حسب ماتدليه عليه عاطفته بحيث لا يمكن تفسير تصرفاته بطريقة منطقية؛

- الزبون الرشيد: هو الزبون الذي يتصرف بعقلانية ورشادة وحكمة فهو عكس الزبون العاطفي حيث أنه دائما ما يسعى إلى تحقيق المنفعة جراء أي سلوك؛

- الزبون الودود: هو زبون محبذ لدى البائعين كونه لا يسبب أي مشاكل في نقاط البيع ولا يعيق عملية البيع والشراء بل دائما يسعى إلى دعم الإتصال بينه وبين البائع؛

- الزبون الإنفعالي: يتعامل مع المواقف بإنفعالية فهو يتميز بالمزاج المتقلب وإتخاذ قرارات شراء عشوائية دون أي تفكير؛

- الزبون الهادئ: هو شخص يتميز بالتأني في إتخاذ قراراته وتحليل الخيارات المتاحة بدقة قبل إتخاذ قراره فهو لا يتأثر بالإعلانات بل يعتمد على تقييمه الشخصي؛

- الزبون الخجول: يعتبر هذا النوع مشكلة وعائق بالنسبة للمؤسسة فهو لا يفصح عن رأيه ويصعب توقع سلوكه الشرائي. (كنزة، سكر، 2011/2010، صفحة 48)

ب- على أساس أهمية الزبون بالنسبة للمنظمة:

-الزبون الإستراتيجي: يتميز هذا الزبون بولاءه القوي للمؤسسة و يعتبر من الزبائن الأكثر مردودية حيث أنه يبحث على علاقات شراكة طويلة الأمد؛

- الزبون التكتيكي: بالرغم من أن هذا النوع يعتبر أقل مردودية من سابقه إلا أنه يحتل مكانة ذات أولوية في المؤسسة بسبب إمكانية تحوله إلى زبون إستراتيجي كما أنه يساعد على ثبات و إستقرار إيرادات المؤسسة من خلال عمليات شرائه المتكررة؛

- الزبون الروتيني: " هو زبون يتساوى إحتمال إستمراره في التعامل مع المنظمة مع إحتمال قطع العلاقة، يعرف على أنه الزبون المشكل، بمعنى يمكن أن يمثل للمؤسسة فرصة أو تهديدا" (كنزة، سكر، 2011/2010، صفحة 51/50)

ج- على أساس نوع الإرتباط بين الزبون والمنظمة:

تتمثل العلاقة بين الزبون والمنظمة في عدة نقاط تقاطع بين متغيرين أساسيين وهما حاجة الزبون إلى العلاقات وحاجة المؤسسة إلى المعلومات والشكل التالي يوضح العلاقة التي تجمع بين المؤسسة والزبون

الشكل 1.1: مصفوفة زبائن المنظمة

المصدر: (سكر كنزة ، 2011/2010، ص50)

يوضح الشكل أربعة أوضع مختلفة تمثل العلاقة بين الزبون و المؤسسة بناء على متغيرين أساسين :

- أولا: المحور العمودي ويمثل الحاجة إلى العلاقات حيث كلما زادت حاجة الزبون للعلاقات فإنه يسعى إلى إنشاء علاقة طويلة الأمد مع المؤسسة ويزيد من ولاءه وثقته والعكس صحيح، إذا كانت رغبته في العلاقات منخفضة فإنه لا يهتم بالعلاقات طويلة الأمد بل يهدف الى إتمام المعاملة وفقط؛

- ثانيا: المحور الأفقي يمثل الحاجة إلى المعلومات حيث إذا كانت حاجة الزبون إلى المعلومات حول منتج معين مرتفعة فإنه يسعى إلى معرفة أدق التفاصيل حول المنتجات أو الخدمات قبل إتخاذه لقرار الشراء، أما إذا كانت منخفضة فإنه لا يبحث عن الكثير من المعلومات بل يكتفي بتجربته السابقة أو عوامل أخرى مثل السعر.

د- التصنيف على أساس العائد والتكلفة:

تلعب التكلفة والعائد دورا حاسما في إتخاذ الزبون لقراراته وتفاعله مع المؤسسة بسبب نطرية التبادل الإجتماعي والتي تفترض أن الزبائن يقومون بعملية موازنة بين ما سيحصلون عليه وما سيدفعونه؛

-زبون بتكاليف عالية وعائد مرتفع: يحقق هدا النوع عادة أرباح مرتفعة إلا أنه في سبيل تحقيقها يقوم أيضا بالمخاطرة وإستثمارات كبيرة عادة ما يكونون شركات أو مؤسسات كبرى تسعى المؤسسة إلى الحفاظ عليهم؛

-زبون بتكاليف عالية وعائد منخفض يتطلب هذا النوع موارد كثيرة من قبل المؤسسة إلا أنهم لا يحققون أي عوائد؛

-زبون بعوائدعالية وتكاليف منخفضة: تعتبر هذه الفئة من الزبائن هي المثالية بالنسبة للمؤسسة كونها تحقق عوائد عالية دون الحاجة الى إستثمار ضخم؛

-زبون بتكاليف منخفضة وعائد منخفض: هذه الفئة إن لم تتحسن عوائدهم غالبا ماتقوم المؤسسة بالتخلي عنهم.

وبرغم إختلاف وتعدد التصنيفات إلا أن نجاح المؤسسة يعتمد على الفهم الجيد والعميق لسلوك الزبائن حيث أن كل زبون لديه عوامل تؤثر على قراراته وفهم هذه العوامل يدفع المؤسسة إلى زيادة أرباحها وتحقيق رضا الزبون وكسب ثقته وولائه لها.

رابعا: مميزات السلوك الإستهلاكي للزبون

يتميز السلوك الإستهلاكي للزبون بعدة مميزات من أهمها:

- تنوع وإختلاف السلوك الإستهلاكي من شخص لآخر وهذا راجع إلى عوامل نفسية مثل مشاعر الفرد تجاه منتج أو خدمة معينة أو عوامل إجتماعية مثل العادات والتقاليد أو ثقافة الفرد و كذلك عوامل إقتصادية ترجع إلى الدخل وسعر السلعة أو الخدمة وأيضا العامل الشخصي مثل العمر أو الجنس أو المستوى التعليمي كل هذه العوامل تؤثر على إتخاذ قرار الشراء وتختلف من زبون لآخر؛

- يعتبر السلوك الإستهلاكي للزبون نشاطا يعتمد على العقل والجسد في نفس الوقت فمن ناحية العقل أنه يتيح له المفاضلة بين البدائل والخيارات المتاحة وإختيار الأنسب من بينها، أما من الناحية الجسدية أنه يقوم بالتسوق أو التحدث مع البائعين؛

- يمكن أن يكون سلوك الزبون في بعض الأحيان غير متوقع بسبب تدخل فرد آخر مثل الأصدقاء أو العائلة في إتخاذ القرار وأحيانا أخرى تلعب الضغوطات النفسية أو الخبرة المحدودة تجاه سلعة معينة في عدم إتخاذ القرار المناسب؛

- يعتبر سلوكا هادفا أي يرتبط بتحقيق رغبة ما وإشباع حاجة معينة، حيث أن السلوك لايظهر من فراغ بل يكون سببه مرتبطا بهدف معين؛

- يعتبر صفات الأفراد مرنة أي أنها تتغير حسب الأوضاع المعاشة وتستجيب للتغيرات الحاصلة سواء تغيرات إقتصادية أو إجتماعية.

تساعد دراسة سلوك المستهلك على تحقيق عدة أهداف إستراتيجية مثل فهم الإحتياجات والرغبات الخاصة بالمستهلكين من قبل الشركات مما يساعدها على تحسين وتطوير خدماتها وسلعها و جذب الزبائن وتحقيق رضاهم وتحسين خدماتهم وتلبية حاجياتهم بشكل أفضل وبالتالي تحقيق الهدف الأساسي للمؤسسة الإقتصادية وهو تحقيق أكبر ربح ممكن وتقليل المخاطر وإتخاذ قرارا ت مثلى كما أن دراسة سلوك الزبون وإمكانية التنبؤ به من خلال تحليل البيانات السلوكية يساعد على منح ميزة تنافسية أكبر للمؤسسات ويمكنها من التكيف مع التغيرات في السوق وزيادة كفاءتها من خلال فهم العملاء بشكل أعمق وهذا ما يجعل المؤسسة تحقق الإستدامة المالية.

المطلب الثاني: محددات السلوك الإستهلاكي

يسعى الأفراد بطريقة مستمرة إلى محاولة إشباع رغباتهم المتنوعة والمتعددة والمتجددة وتحقيق أهدافهم، من خلال القيام بعملية الشراء أو تغير توجه الفرد بإقتنائه لعلامة تجارية معينة دون أخرى، قد يظهر للوهلة الأولى أن سلوك الفرد المستهلك مجرد تصرف عشوائي وغير رشيد إلا أنه يمكن أن يكون

صادرا عن تفكير طويل ودراسة شاملة ومتأنية للعلامات التجارية البديلة والمتاحة والمتوفرة في السوق، لهذا فإن إعتبارنا لسلوك الزبون أنه مجرد تصرف غير مدروس أو مخطط له فإنه حكم ربما يكون متسرعا، حيث أن سلوكه تتحكم فيه عدة عوامل ومحددات تتمثل في:

أولا: العوامل النفسية

أ- الدافع: يلعب دورا أساسيا في إتخاذ القرار ويقصد بالدافع هو الرغبة التي تقود الفرد وتوجه سلوكه نحو إتخاذ قرار معين، كما يمكن تعريفه على أنه القوة الكامنة أو الداخلية للفرد ويمكن أن يكون الدافع داخليا مثل الرغبة في التعلم أو تحقيق الذات أو يكون خارجيا مثل القيام بعمل معين بهدف الحصول على مكافآت مادية أو جوائز؛

كما يعتبر الدافع نقطة الإنطلاق التي يمكن من خلالها معرفة وفهم سلوك المستهلك ويمكن تعريف الدافع أيضا على أنه "حالة من التوتر الداخلي الناتج عن نقص في حاجة من حاجات الفرد شأن هذه الحاجة أن تدفع الفرد للقيام بسلوك خاص صوب إشباع الحاجة" (سعاد، 2023، صفحة 17/16).

ب- الحاجات: تبدأ عند الإنتباه إلى وجود حاجة غير مشبعة فهي حالة يشعر فيها الفرد بالنقص أو الحرمان تجاه أمر معين حيث يدفعه هذا الشعور إلى البحث عن طريقة لإشباعه، بحيث يمكن أن تكون الحاجة فطرية(أساسية) مثل الطعام، الماء، النوم،أو الراحة، التقدير،...إلخ أومكتسبة (ثانوية)، كما يمكن تعريف الحاجة على أنها حالة من الإضطراب أو التوتر الذي ينشأ بسبب عدم إشباع رغبات الفرد، وقد تطرق أبراهام ماسلو في نظريته التي قدمها في ورقته البحثية نظرية الدافع البشري سنة 1943، والتي تهتم بدراسة تدرج حاجات الأفراد مثلها في شكل هرم ووضح فيها ترتيب حاجات الأفراد الخمس حسب أولوياتها بدءا بالقاعدة التي تمثل

-الحاجات الفيسيولوجية: والمتمثلة في المتطلبات الأساسية بكل فرد مثل الطعام، الماء، الهواء، النوم،..إلخ؛

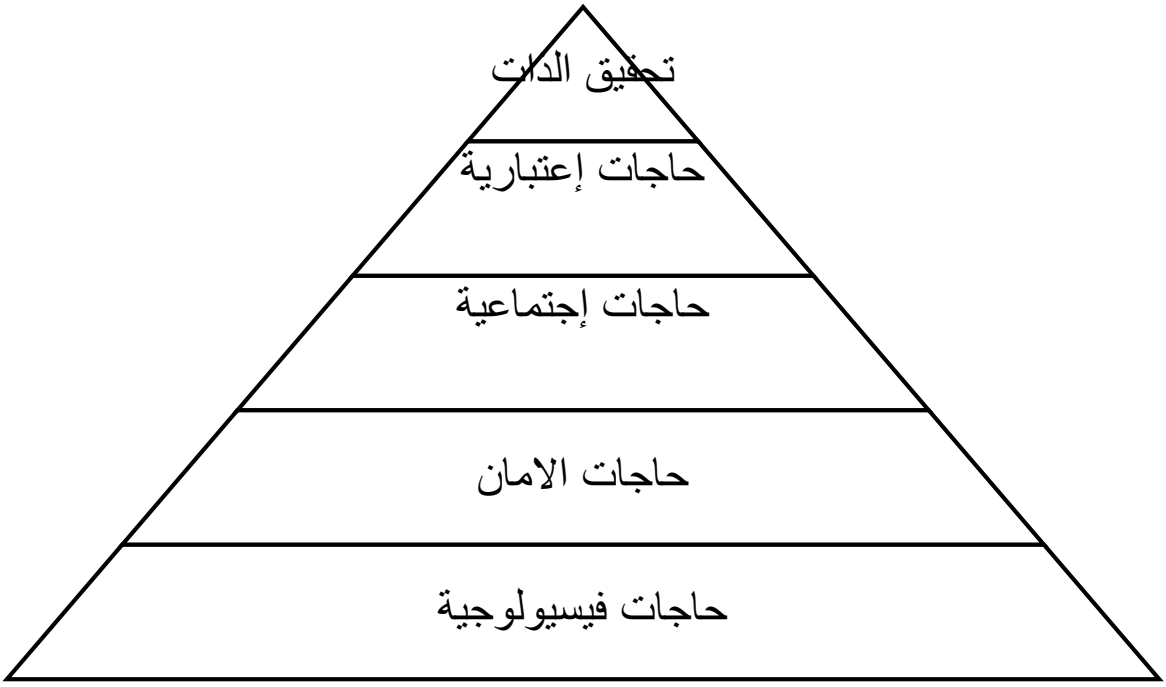
-الحاجة إلى الأمان والإستقرار والإنتماء والرغبة في التواصل مع الأخرين؛

-الحاجات الإجتماعية مثل إحتياج الفرد إلى المجتمع الذي ينتمني إليه؛

-الحاجات الإعتبارية أو الحاجة للتقدير ويقصد بها الرغبة في الحصول على الهيبة والمكانة الاجتماعية والثقة والشعور بالفخر؛

- تحقيق الذات في هذه المرحلة يصل الفرد إلى تحقيق ذاته والوصول إلى مراده من خلال إنجازاته والقيام بأنشطة إبداعية وتحقيق طموحاته ومآربه الشخصية. (سعاد، 2023، صفحة 18/17)

الشكل 2.1: هرم موسلو للحاجات



المصدر: (بورحلة سعاد، 2023، ص 18)

ج- الإدراك: ويقصد به الإعتقاد الذي يكونه الفرد حول أمر ما ويعرف مصطلح الإدراك على أنه"العمليات العقلية المتظمنة في إكتساب المعرفة والفهم" (إستعراف، 2025) فإعتقاد الشخص مثلا أن الوظيفة لاتناسبه أو أن جو عمله غير مريح أو ملائم يقوم بتحديد طريقة تصرفه في العمل وتحديد سلوكه بغض النظر عن الواقع حيث أن الإنسان وحسب فطرته يستجيب مع شخص يعرفه أو يملك تجاهه إنطباع جيد أو إيجابي أفضل من إستجابته لشخص آخر لايملك تجاهه أي إعتقاد أو يكون إنطباعه سيء نحوه ونفس الأمر ينطبق على تعامل الزبون المستهلك مع السلع والخدمات، حيث يساعد إدراك الفرد المستهلك بصفة كبيرة على نجاح أو فشل السياسات التسويقية من خلال الشعور أو النظرة التي يملكها الزبون حول المنتج والتي تدفعه لإتخاذ قرار الشراء من عدمه أو لمواصلة تعامله مع المؤسسة أو تخليه عنها وإستبدالها؛

د- التعلم: وهو عملية يقوم بها الفرد بهدف إكتساب معلومات أو خبرات أو مهارات جديدة مما يدفعه إلى تغيير سلوكه أو أداءه أو تفكيره تجاه موضوع ما أو فكرة معينة، بحيث يشمل التعلم عدة جوانب منها نفسية،

إجتماعية، ثقافية،..إلخ، وينقسم التعلم إلى3 أنواع تعكس إختلاف طرق فهم الأفراد وإستيعابهم للمعلومات وهي:

- التعلم الفكري: وهو عملية إستيعاب المعلومات من خلال القراءة، البحث، الخرائط الذهنية، الإستماع، المشاركة في المناقشات وذلك بهدف تحسين القدرة على التفكير والتحليل وتحسين الأداء؛
- التعلم بالملاحظة: أو التعلم بالمشاهدة أي يمكن أن يكون الفرد نمط من خلال مشاهدة أو ملاحظة كيف يتصرف الآخرون تجاه مشكلة معينة وطريقة تعاملهم معها؛
- التعلم بالتسلسل الفكري: و ذلك بالإعتماد على قدرة التفكير والتحليل والربط بين المعلومات؛
- التعلم بالإرتباط: يقصد به "كيفية تذكر المستهلكين للكلمات يجب أن تكون مرتبطة ببعضها وقد سميت المهمة التي تتطلب الأفراد ربط سلسلة من الإستجابة المزايا المرتبطة بالمنتج" (هدى، 2020/2019، صفحة 74).

ثانيا: العوامل الإقتصادية

الدخل: حسب النظرية الإقتصادية فإن الدخل والإستهلاك تربط بينهما علاقة طردية بشكل عام وهذا راجع إلى نوع السلعة أي أنه كلما زاد الدخل سيزيد الإستهلاك فعندما يكون دخل المستهلك محدودا فهذا يجعل المستهلك يفكر في شراء السلع الأساسية فقط والتي يجب شراءها مثل الأكل و الماء، وعادة السلع الأساسية لا تؤثر على الدخل بشكل كبير فعند إنخفاض الدخل يلجأ المستهلك إلى السلع البديلة مثل إستبدال اللحوم الحمراء باللحوم البيضاء، بينما إذا إرتفع دخله فإنه يتغير تفكيره إلى شراء السلع الفاخرة مثل المجوهرات والسيارات،... ويمكن القول أن الدخل له علاقة جوهورية بالإستهلاك؛

السعر: وهو القيمة النقدية التي يدفعها المستهلك للحصول على سلعة أو خدمة معينة، بحيث يعتبر السعر عاملا إقتصاديا مهما في التأثير على السلوك الإستهلاكي للزبون بصفة مباشرة بحيث دائما ما يسعى الزبون المستهلك إلى الحصول على السلعة أو الخدمة بأقل تكلفة ممكنة وتتحدد علاقة السعر بالمستهلك حسب نوعي السلعة و تكون العلاقة بين السعر والطلب علاقة عكسية فكلما زاد سعر السلعة نقص الطلب عليها إلا أنه هناك بعض الإستثناءات المتمثلة في:

-السلع العادية مثلا(القهوة) عندما يرتفع سعرها فإن المستهلك يلجأ للبحث عن بديل أرخص مثل (الشاي)؛

-السلع الضرورية مثلا (الخبز، الماء، الأدوية) هذه السلع يحتاجها المستهلك بإستمرار ولا يمكنه الإستغناء عنها لهذا فإنه يستمر في الطلب عليها حتى مع إرتفاع سعرها؛

-السلع المكملة أو الفاخرة مثل(الجواهر ، الساعات الغالية، السيارات،...) أحيانا يزيد الطلب عليها حتى مع إرتفاع سعرها لأن الزبون المستهلك دائما مايربط الجودة بالسعر ويؤمن بأنه كلما زاد السعر تحسنت الجودة؛

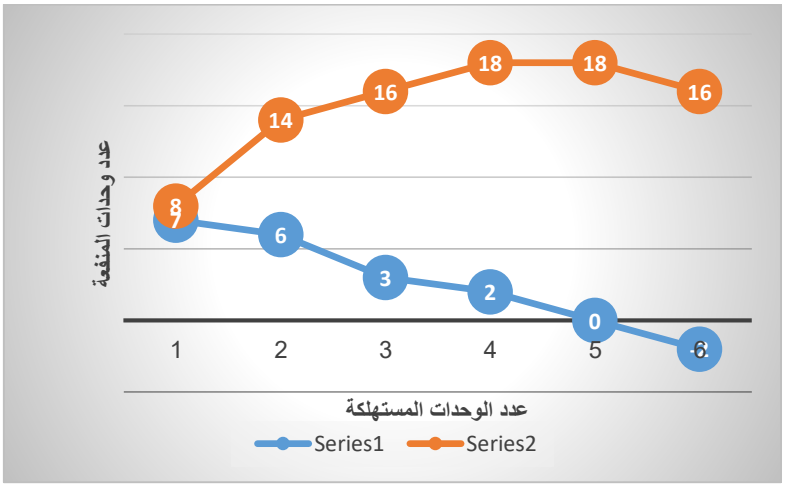
المنفعة: وهي قابلية أو قدرة إشباع السلعة أو الخدمة حاجة المستهلك خلال فترة زمنية محددة كما تعرف بأنها مقدار الإشباع الذي يحصل عليه الفرد عند إستهلاكه لسلعة معينة بحيث دائما مايسعى المستهلك لتحقيق أقصى منفعة في حدود دخله، وتنقسم المنفعة إلى نوعين:

أ –منفعة كلية(Total Utility): وهي مجموع المنافع التي يقوم الفرد بإستهلاكها من نفس السلعة من خلال وحدات متتالية في فترة زمنية ما؛

ب –منفعة حدية(Marginal Utility): وهو التغير الذي يحصل في المنفعة الكلية بسبب التغير في كمية إستهلاك السلعة بوحدة واحدة في فترة زمنية ماوتقوم نظرية المنفعة على عدة إفتراضات من بينها:

- قابلية القياس كميا بحيث يمكن إعطاء قيمة رقمية للمنفعة وهذا مايساعد المستهلك في ترتيب تفضيلاته حسب الأولية وإتخاذ قراراته الإستهلاكية؛
- منفعة المستهلك من سلعة ما مستقلة أي لا تتأثر بمنفعته من سلعة أخرى أي يمكن للمستهلك تحقيق أقصى إشباع ممكن؛
- تتميز المنفعة الحدية بالتناقص بحيث كلما إرتفع مستوى إستهلاك سلعة معينة إنخفضت المنفعة.

الشكل3.1: منحنى المنفعة الحدية والمنفعة الكلية



المصدر:(معاذ الشرفاوي ، 2020، ص69)

يوضح المنحنى السابق العلاقة بين المنفعة الكلية(اللون البرتقالي) والمنفعة الحدية(اللون الأزرق) مع عدد الوحدات المستهلكة من سلعة معينة

نلاحظ أن المنفعة الكلية تكون متزايدة في البداية ثم تصل إلى أقصى حد لها عند الوحدة الخامسة(18 منفعة)، أما عند الوحدة السادسة تنخفض المنفعة الكلية إلى (16 منفعة) وهذا يعكس شعور المستهلك بعدم الرضا أي أن قيمة الإشباع أصبحت زائدة عن الحاجة؛ حسب المنحنى فإن المنفعة الحدية تكون موجبة في البداية لكنها تبدأ بالتناقص تدريجيا بسبب قانون تناقص المنفعة الحدية يعني أن كل وحدة إشباع تضيف منفعة أقل من الوحدة التي قبلها، عند وصول المنفعة الحدية إلى الصفر تحديدا عند الوحدة الخامسة يعني أن المستهلك لم يعد يحصل على أي إضافة أو زيادة في الإشباع من إستهلاك أي وحدة إضافية، بعدها تصبح المنفعة الحدية سالبة وهو مايعكس إنخفاض المنفعة الكلية وشعور المستهلك بالإشباع الزائد وعدم الرغبة في إستهلاك المزيد؛

يفضل الزبون في هذه الحالة التوقف عن الإستهلاك عند وصول المنفعة إلى أقصاها بسبب أن إستمرار إستهلاكه سيؤدي إلى تقليل المنفعة الإجمالية.

المطلب الثالث: سلوكيات الزبائن الممكن التنبؤ بها

إن السلوكيات التي يمكن أن يظهرها الزبون أثناء تعامله مع المنتجات أو الخدمات متعددة وتتغير حسب إحتياجه وتجاربه وهي ما تحدد تصرف الزبون وتدفعه لإتخاذ القرارات ومن أهم هذه السلوكيات هي:

أولا: السلوك الشرائي

يعتبر قرار الشراء هو ماتطمح إليه المؤسسات وتنتظره من الزبون والتنبؤ به من خلال تحليلها لأنماط الشراء الخاصة بالزبائن وكذلك العوامل التي تؤثر على قرار الشراء مثل الجودة، السعر،...إلخ يساعد المؤسسة على معرفة التوقيت المثالي للتسويق والترويج لمنتجاتها بهدف تحقيق أكبر عائد ممكن، وتمر عملية إتخاذ القرار الشرائي بعدة مراحل وهي:

المرحلة الأولى: مرحلة ما قبل الشراء

تبدأ هذه المرحلة عند شعور الزبون بالحاجة إلى منتج معين والرغبة في الحصول عليه، وهذا ما يدفعه للبحث عن معلومات حول المنتج ومن ثم تقييم البدائل المتاحة وإختيار أفضل بديل؛

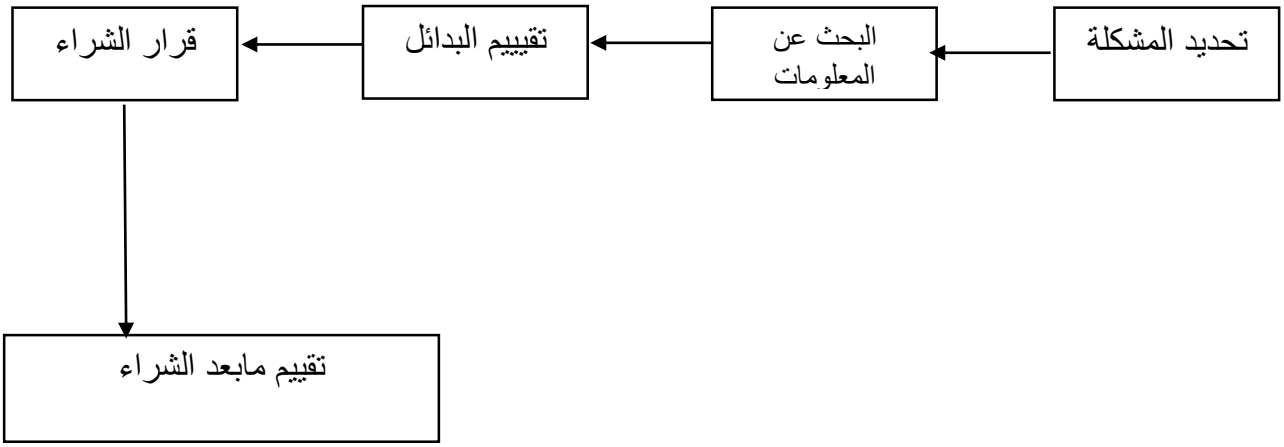
المرحلة الثانية: مرحلة الشراء؛

المرحلة الثالثة: مرحلة ما بعد الشراء؛

وفيها تتم عملية الشراء الفعلية سواءا بالإعتماد على إستراتيجية محددة إعتاد عليها أو إتخاذ إستراتيجية جديدة و يقوم الزبون بإستهلاك المنتج وتقييمه. (الجريسي، 1427هـ، صفحة 86)

و يمكن توضيح كيفية إتخاذ القرار الشرائي من قبل الزبون من خلال الشكل التالي:

الشكل4.1 : مراحل إتخاذ القرار الشرائي



المصدر: (Patricia cautelle, Marketing delanalyse, node et Pearson education, Paris, 2004,p105.)

كما يمكن التنبؤ بالطلب على المنتجات التي تكون متلازمة أي يتم شراؤها في نفس الوقت مثل السلع المكملة وذلك من خلال تحليل سلة التسوق وذلك بهدف تجنب العجز أو الفائض في الإنتاج.

ثانيا: ولاء الزبون للعلامة التجارية

يعد إلتزام الزبون تجاه العلامة التجارية أمرا لا بد منه في المساهمة في نجاح المؤسسة إلا أن هذا الولاء لا يأتي من فراغ بل بسبب عنصر جوهري وهو رضا الزبون حول المنتج أو الخدمة التي تقدمها تلك المؤسسة حيث أن العمل على إرضاء الزبون واجب على المؤسسة كونه يقدم مزايا مهمة سواءا للزبون أو المؤسسة مثل الميزة التنافسية التي يقدمها للمؤسسة كما أنه يمنح نوعا من الإستقرار لبرامج وخطط العمل خلال فترة معينة من الزمن، كما تدفعه إلى تكرار عملية الشراء مما يؤدي إلى إلتزامه لتلك العلامة، ويبعث الرضا أيضا شعور بالثبات والإستقرار في نفس الزبون ويخلق له حافزا وسببا لمواصلة التعامل مع نفس العلامة، و حتى تتمكن المؤسسة من كسبه والحفاظ عليه لأطول مدة ممكنة عليها أن تعرف محددات الرضا لدى الزبون والمتمثلة في:

أ – التوقعات: أو ما يعرف بالقيمة المتوقعة وهو الأداء أو النتيجة الذي ينتظرها الزبون والتي يخطط لها ويتوقع حدوثها من قبل المؤسسة قبل القيام بعملية شراء المنتج، وتتعلق هذه التوقعات بعدة جوانب وهي:

-خصائص ومميزات المنتج أو الجودة التي يتميز بها عن باقي المنتجات الأخرى المنافسة؛

-المنفعة التي سيحصل عليها الزبون جراء إقتناءه لهذا المنتج وهل تتوافق هذه المنفعة مع رغبته؛

-تكلفة المنتج هل ستكون مناسبة لدخله ويتمكن من تنفيذ الشراء.

ب – إدراك الأداء: تأتي هذه المرحلة بعد شراء الزبون للمنتج وتسمح له بمعرفة وإدراك نسبة توقعاته وهل كانت تجربته مع هذا المنتج إيجابية وتحققت توقعاته أو العكس، ويمكن للزبون تقييم المنتج وأدائه سواء كان مثالي، متوسط، سيئ.

ت – المطابقة : أو المقارنة بين مستوى أداء المنتج وما كان يتوقعه الزبون ويطمح إليه، حيث أنه كلما كان التوقع أقرب إلى أداء المنتج زادت إحتمالية ولاء الزبون (نجود، صفحة 240/239)

يمكن للمؤسسات أن تتنبأ بولاء الزبائن من خلال سلوكهم الشرائي المتكرر وكذلك تفاعلهم الإيجابي مع العلامة وهذا ما يحسن من القررات التي تتخذها المؤسسة ويزيد من فرصة الإحتفاظ بزبائنها لأطول مدة ممكنة.

ثالثا: طريقة إتخاذ القرار

يمكن التنبؤ بكيفية إتخاذ قرار الزبون من خلال معرفة التركيبة السلوكية أي كيف يتصرف الزبون والتركيبة الديموغرافية من خلال إستراتيجية معينة تعرف بالتسويق المستهدف أو إستهداف السوق، والتسويق هو طريقة ذكية تستطيع المؤسسة الإقتصادية إستعماله بهدف تحقيق أكبر ربح ممكن و الوصول إلى الحرية المالية، و يمكن تعريف التسويق بأنه" وسيلة لزيادة العوائد التي تحصل عليها من عمل تجاري وتلك الفوائد توفر لك دخلا يجعلك تعيش حياة أفضل" (شبايك، 2009، صفحة 5)، أما التسويق المستهدف فهو عبارة على إستراتيجية يمكن من خلالها معرفة الفئة أو الشريحة التي تمتلك أكبر نسبة ملائمة وتناسب مع المنتج المراد التسويق له وذلك من خلال خطوات تتمثل في:

-تقسيم السوق Segmentation: أي تجزئة العملاء وتقسيمهم بناءا على معايير ديموغرافية وإقتصادية مثل (العمر، النوع، الدخل، الإهتمامات، المستوى التعليمي، ...إلخ)، ويمكن من خلال هذه البيانات معرفة الفئة التي يمكن أن تحقق أكبر عائد مالي وكذلك تكون هذه الفئة قابلة للزيادة مع مرور الوقت مع الحرص على توفير ميزات مثل التوصيل خاصة للفئة العاملة التي لاتملك وقتا. (شبايك، 2009، صفحة 12)

من خلال التسويق المستهدف تستطيع الشركات التنبؤ بتفاعل الزبائن مع حملات الترويج وطريقة إستجاباتهم للتسويق، ومعرفة الوقت المناسب للشراء وكذلك قنوات الشراء المفضلة لديه فمثلا شركات

مستحضرات التجميل تستهدف النساء الذين يهتمون بالعناية بالبشرة أما الشركات مثل NIKE وAdidas تستهدف فئة الرياضيين والذين يميلون للياقة البدنية وفكرة تقسيم السوق تساعد على إختصار طريق النجاح ومعرفة الفئة المناسبة بصفة مباشرة لتوفير الوقت والتكاليف.

رابعا: مغادرة الزبون

يمكن للمؤسسة الحفاظ على زبائنها و تفادي مغادرتهم من خلال العمل على تحسين تجربتهم وكسبهم وتعرف تجربة الزبون على أنها عبارة على تراكم لشعورهم عند كل تعامل مع علامة تجارية معينة، وتلعب خدمة الزبائن أمرا هاما وأساسيا في بناء سمعة المنتج و تعزيز الإنطباع الجيد لدى الزبون بحيث يمكن من مجرد معرفة المؤسسة كيف تعامل زبائنها يدفع الزبائن إلى شراء المنتج وبالتالي كسب ولاء الزبون خاصة وقد أصبحت الشركات في تنافس مستمر وكل واحدة تسعى إلى كسب أكبر عدد ممكن من العملاء، ومن أجل تحسين تجربة العملاء يمكن إستخدام بعض الإستراتيجيات مثل:

-الإستثمار في العلامة التجارية والعمل دائما على تحسينها وتقريبها من ما يريده الزبون؛

-إستخدام روبوتات الدردشة مع الزبائن لمعرفة إهتماماتهم ومايروق لهم في المنتج والعكس، كما أن طريقة الدردشة تعتبر فعالة وموفرة للوقت مقارنة بالطرق التقليدية مثل الإستبيانات (Hesham Moussa, n.d.)إضافة إلى أنه يمكن التنبؤ بشعور الزبائن تجاه منتج معين من خلال تحليل تعليقاتهم وآرائهم على منصات التواصل الإجتماعي، بهذه الطريقة تستطيع الشركات ضمان الزبائن بنسبة أكبر بإعتبار أن سلوك الزبون المتمثل في المغادرة يعتمد بشكل كبير على تجربته مع الشركة ومدى شعوره بالرضا.

المبحث الثاني: طرق التنبؤ بإستخدام التعلم الآلي

سعى الإنسان منذ القدم على التطلع إلى المستقبل ومعرفة ما يحدث من حوله من تغيرات في الطبيعة والظواهر المحيطة به، وهذا راجع إلى طبيعته الفضولية حيث كان يلجأ إلى إجراء عملية التنبؤ بالإعتماد على معرفته وخبرته وحدسه، فقد كان البشر قديما يستعملون علم التنجيم وقراءة الطالع ورصد حركة النجوم للتنبؤ بالأحوال الجوية كما كان الصيادون يحتاجون إلى معرفة مكان الفرائس وهذا يدل على قيمة وأهمية التنبؤ ومع مرور الوقت تطورت وسائل التنبؤ من مجرد تخمينات قائمة على الحدس والشعور إلى أساليب علمية وعملية تعتمد على البيانات والإحصائيات ونطرا لأهمية التنبؤ في العديد من المجالات و التي من بينها الإقتصاد وخاصة في عالم أصبح يضج بالكثير من البيانات و التي يجب تحليلها وإستنتاج

المعلومات منها بغية توقع ما سيؤول إليه الوضع في المستقبل حييث توجد عدة طرق للتنبؤ لكن لا يمكن حصرها في مبحث واحد، لذا سيتم التطرق فقط إلى التقنيات التي لها علاقة مباشرة بهذه الدراسة.

المطلب الأول: عموميات حول التنبؤ

أولا: تعريف التنبؤ

يمكن تعريف التنبؤ على أنه توقع ما سيحدث في المستقبل لظاهرة معينة بناءا على إتجاه هذه الظاهرة وسلوكها في الماضي بناءا على نماذج التنبؤ المعروفة (العبيد، 2003، صفحة 02) و التنبؤ هو عملية تخطيط و توقع أوتقدير لما سيؤول عليه المستقبل بناءا على بيانات سابقة أو حالية، بهدف الإستعداد للمستقبل وإتخاذ القرارات المناسبة وإستغلال الفرص بشكل أمثل وتجنب المخاطر المحتملة، دائما ما يعتمد الأفراد أو الحكومات على التنبؤ من أجل إتخاذ قرارات إستراتيجية والحد من المخاطر.

ثانيا: أنواع التنبؤ

أ- حسب طول الفترة الزمنية

1. التنبؤ قصير المدى: يستعمل هذا النوع في حالة قصر الفترة الزمنية عادة ما تكون الفترة أقل من 3 أشهرحيث تجعله هذه الفترة مناسبا أكثر للمؤسسات كونه ويتميز بنتائج عالية الدقة بسبب أن التغيرات التي تحدث في هذه الفترة الصغيرة يمكن توقعها بسهولة مقارنة بالفترة طويلة الأجل وهذا مايدفع المؤسسة لتحسين موقعها التنافسي.

2. التنبؤ متوسط المدى: تمتد فترة التنبؤ في هذا النوع من 3 أشهر إلى 3 سنوات وتبرز أهمية التنبؤ متوسط المدى في إمكانية التوسع في مشروع أو صناعة معينة بغرض التخطيط للإنتاج أو المبيعات وتحليل مختلف الخطط التشغيلية.

3. التنبؤ طويل المدى: تزيد الفترة فيه عن خمس سنوات، ويستعمل في التخطيط الإستراتيجي والمالي وإدارة المخاطر، إلا أنه يعتبر أقل دقة بسبب طول الفترة حيث تعلو فيه درجة عدم اليقين (جامعة محمد بوضياف، 2005/2006، صفحة 09)

ب- حسب طبيعة البيانات:

1. التنبؤ الكمي: يعتمد على بيانات رقمية مثل التنبؤ بالكميات المطلوبة من منتج معين، أو التنبؤ بمؤشرات النمو الإقتصادي مثل التنبؤ بالناتج المحلي الإجمالي أو معدل التضخم...إلخ.

2. التنبؤ النوعي: تكون البيانات المستخدمة في هذا النوع بيانات نوعية ترجع إلى الخبرة وتوقعات الخبراء في عدة إختصاصات أو مجالات مثل الموضة أو السياسة أو السوق وغيرها.

ت- حسب أسلوب التنبؤ

1. التنبؤ الإستنتاجي: يقوم هذا النوع على فرضيات، ونظريات علمية كما يستند على تحليل العلاقات السببية بين المتغيرات؛

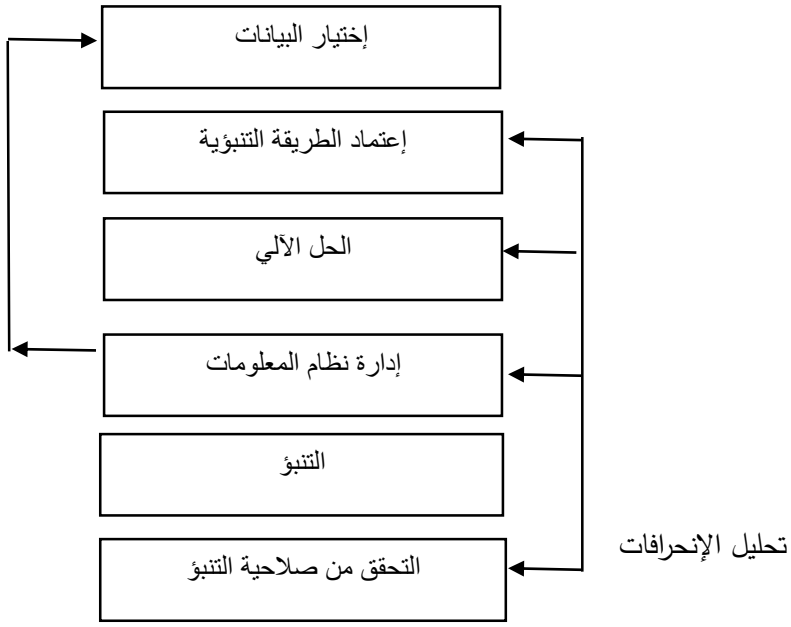
2. التنبؤ الإستقرائي : يعتمد على تحليل البيانات التاريخية مثل التنبؤ بأرباح المنشأة بناءا على الأداء المالي السابق؛

3. التنبؤ الحدسي: يعتمد على التجارب السابقة والحدس دون اللجوء إلى بيانات كمية واضحة.

ثالثا: مراحل التنبؤ

تمر عملية إجراء التنبؤ بعدة مراحل موضحة في الشكل التالي:

الشكل 5.1: مراحل التنبؤ



المصدر: (ريجي بوربوني ، جان كلود إيزينيه، 2008 ، ص(25))

-إختيار البيانات: ويقصد بها البيانات المتوفرة والواجب إختيارها للتنبؤ بالظاهرة المدروسة والتي تعتبر نقطة الإنطلاق لعملية التنبؤ، بحيث يجب أن تكون عدد المشاهدات كافية؛

-تردد البيانات periodicity: وهي الفترة الزمنية الفاصلة بين المشاهدات حيث أن هذه الفترة يحددها نوع نشاط المنشأة ونوع المنتج؛

- إختيار الطريقة المناسبة: توجد تقنيتان لمعرفة الطريقة المناسبة وهما:

التقنية الأولى: تعتمد على طرائق التمديد الخارجي extrapolates فكرتها الأساسية هي عزل عدد محدد من مركبات السلسلة وتمديد السلسلة المتبقية تجاه المستقبل المراد التنبؤ به؛

التقنية الثانية: وهي الطريقة التفسيريةexplicative تهدف للبحث عن علاقة بين السلسة المتنبأ بها وسلسلة أخرى تكون تفسيرية.

كما يمكن الخلط بين التقنيتين لخدمة الدراسة

- الحل الآلي ونظام المعلومات: ويقصد بها إستخدام برامج متقدمة لإتمام عملية التنبؤ مما يوفر الوقت والدقة والسرعة من أجل إخاذ قرارات مبنية على بيانات موثوقة ودقيقة؛

- التنبؤ ومدى صلاحيته: يتم إنشاء النموذج الرياضي للتنبؤ بالإعتماد على المراحل السابقة. (إيزينيه و ريجي بولوني، 2008، صفحة 27).

رابعا: أهمية التنبؤ:

-يساعد التنبؤ على التخطيط المستقبلي وتحديد مسار التطور والتقدم في مختلف المجالات؛

-تحديد الأولويات البحثية ومساعدة العلماء والباحثين على معرفة المجالات التي تحتاج إلى دراسة أو تحسين؛

-الإستعداد للتحديات المتوقعة مستقبلا ومحاولة التكيف مع التغيرات؛

-تحقيق التنمية المستدامة وضمان إستمرارية توفر الموارد للأجيال اللاحقة؛ (صالح، 1978، صفحة 11)

-إستخدام التنبؤ لفترة من الزمن يؤدي إلى تقليص الشك وعدم اليقين في المستقبل إذ يمكن التعبير عنه بطريقة إحصائية بإنخفاض الإنحراف المعياري لخطأ التنبؤ؛

-يمكن تحسين جودة التنبؤ النهائية من خلال ربط نتيجة التنبؤ ببصيرة المتنبئ. (إيزينيه و ريجي بولوني، 2008، صفحة 36)

المطلب الثاني: مدخل إلى تعلم الألة

تعتبر المعلومات في عصرنا الحالي هي المحرك الأساسي للتكنولوجيا والإبتكار و وقودا للشركات كونها تساعد على إكتشاف العلاقة بين المتغيرات وتحسن من أداء النماذج عن طريق تحديثها بإستمرار،

حيث تسعى المؤسسات إلى فهم طريقة إستخدام البيانات لصالحها وتعزيز قدرتها على التنبؤ و إتخاذ القرارات الصحيحة والمستنيرة ومن بين أكثر الأدوات المستخدمة في تحليل البيانات هو التعلم الآلي فهو يركز على تطوير أنظمة لمعالجة البيانات وتحليلها.

1- تعريف تعلم الآلة: هو "مهارة أساسية. تتمثل الفكرة الأساسية للتعلم الآلي في السماح للآلات بالتعلم بشكل مستقل بإستخدام كتلة البيانات التي يتم تغذيتها بالجهاز كمدخلات و مع تقدم التكنولوجيا، يتم تدريب الآلات على التصرف مثل البشر في القدرة على إتخاذ القرار" (وازن، تعلم الآلة وعلم البيانات، صفحة 36) أي هو إستخدام الآلة في إتخاذ القرارات بطريقة ذكية.

2- مكونات تعلم الآلة: يحتاج كل نظام تعلم يستخدم تعلم الآلة إلى 3مكونات أساسية وهي:

أ- البيانات data: تعتبر عنصرا أساسيا في تعلم الآلة يمكن أن تكون البيانات على شكل قيم، نصوص، صور،... خام يتم تحليلها ومعالجتها وتحويلها من مجرد بيانات مشتتة وغير مفهومة إلى معلومات مفيدة يمكن إستخدامها لإتخاذ لإتخاذ القرارات حيث يمكن الحصول عليها إما يدويا أو آليا، فمارك زوكربرغ مثلا بارع جدا في معالجة وتحليل بيانات المستخدمين و أنشطتهم على فيسبوك بهدف معرفة تفضيلاتهم، كما أن شركة google تدعي أنها تتحقق من كونك بشري أو robot إلا أنها في الواقع تستعملك لمساعدتها في تصنيف بياناتها من خلال التحقق البشري captcha والتي تستخدمها في أغلب المواقع، أحيانا يكون الحد الأدنى من البيانات المطلوبة هوعشرات الآلاف لإستنتاج معلومة وأحيانا أخرى نحتاج إلى ملايين العينات؛

ب- الميزات features: أو ما يعرف ب parameters أو variable يمكن أن تكون الميزة عبارة على مسافة أو بيانات خاصة بأفراد أو مستخدمين مثل الإسم، العمر، النوع،...إلخ أو سعر سهم أو كلمات متكررة في نص معين، حيث يجب مراعاة هذه الخصائص عند معالجة البيانات ويكون الأمر أوضح وأبسط في حالة ما كانت البيانات أو الميزات منظمة في جدول؛

ت- الخوارزميات: تعتبر أسهل طريقة لبناء نموذج تعلم آلي هو إختيار الخوارزمية المناسبة وتتحدد من خلال دقة وسرعة النموذج المطلوب، فإذا كانت البيانات بسيطة ليس من المنطقي إستخدام خوارزمية معقدة فالخوارزميات هي الأساس الذي يعتمد عليه نظام التعلم الآلي ويتحدد حسب الظاهرة المدروسة ونوع البيانات وطبيعتها بإعتبار أن الإختيار الصحيح للخوارزمية سيؤثر بشكل كبير على دقة وكفاة النموذج. (لحلح، 2020، صفحة 65/61)

3- أنواع التعلم الآلي: ينقسم التعلم الآلي إلى 3 أنواع أساسية وهي:

أ- التعلم بالإشراف: هو أسلوب تعليمي يستند على وجود مشرف يقوم بالتعليم والتوجيه ويقصد به طريقة إنشاء نموذج للتنبؤ من خلال إستخدام بيانات مصنفة أي أنه كل عينة موجودة في البيانات تكون لها سمات ومميزات تختلف عن باقي العينات فمثلا لفهم المؤسسة لخسارة عملائها يتوجب عليها معرفة الأسباب التي تدفع الزبون إلى الخروج مثل متوسط عملية الشراء، الأيام المنقضية منذ آخر عملية شراء، هل لا يزال الشخص عميلا أو لا من خلال البيانات الموجودة؟...إلخ

كما يمكن تعريفه على أنه عملية تدريب training الآلةعلى بيانات معينة (مدخلات) لإعطاء قيمة (مخرجات) ففي حالة تغير المدخلات تكون الآلة قادرة على إعطاء قيمة جديدة للمخرجات وهناك نوعين رئيسين لطريقة التعلم الموجه أو التعلم بالإشراف وهما:

- التصنيف(Classification): يهدف إلى التنبؤ بصنف معين من خلال تقسيم البيانات إلى فئات أو عناصر حسب ميزة أو سمة معينة مثل تصنيف الملابس حسب اللون، أو تصنيف المستندات حسب اللغة أو حسب الموضوع ....إلخ، وعادة ما يتم إستخدام التصنيف في عدة مجالات مثل:

-تصفية البريد من الرسائل المزعجة؛

-التعرف على الحروف المكتوبة باليد؛

-التنبؤ بتخلي الزبون عن الخدمة؛

-تحليل المشاعر؛

-الكشف عن الغش؛

-البحث عن وثائق مماثلة.

ومن بين الخوارزميات الأكثر إستخداما في عملية التصنيف هي:

- خوارزمية شجرة القرار Decision tree؛
- خوارزمية بايز Naïve Bayes؛
- الإنحدار اللوجستي ؛ Logistic Regression؛
- خوارزمية الدعم الآلي المتجه Support vector machine ؛
- خوارزمية الجار الأقرب K- Nearest Neighbors. (لحلح، 2020، صفحة 69)

- الإنحدار: وهو طريقة لرسم خط يجمع بين عدة نقاط و يتم إستخدام الإنحدار في العديد من المجالات مثل

-الإقتصاد و في البورصة تحديدا حيث يساعد الإنحدار على معرفة أسعار الأسهم، ويمكن أيضا أن يحلل حجم الطلب والمبيعات؛

-التشخيص الطبي من خلال تحليل العلاقات بين الأعراض ونتائج الفحوصات وإحتمال الإصابة بمرض معين.

ومن الخوارزميات الشائعة الخاصة بالإنحدار هي:

- خوارزمية الإنحدار الخطي Linear Regression؛
- خوارزمية الإنحدار متعدد الحواف Polynomial ؛

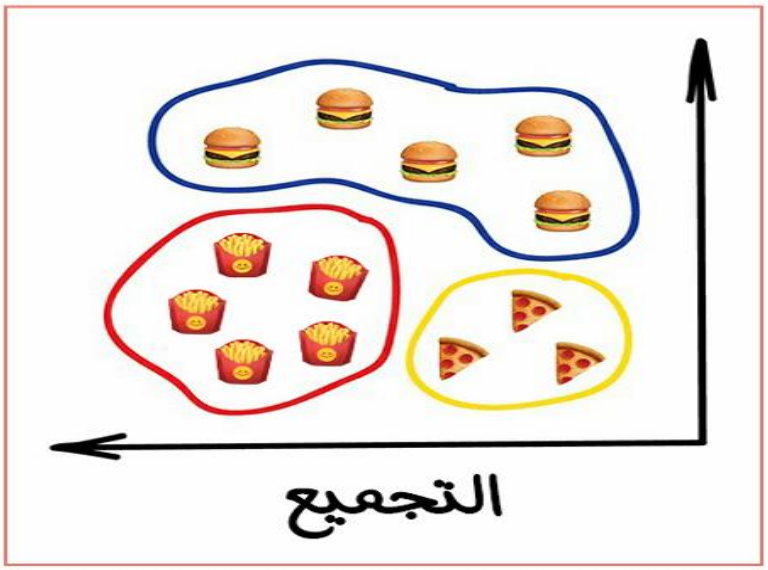
يختلف الإنحدار عن التصنيف بأنه يتنبأ بالقيم والأرقام بدلا من الفئات حيث تحاول الآلة رسم خط الإنحدار بدقة عالية عكس الشخص الطبيعي الذي يقوم بالرسم يدويا وينقسم الإنحدار إلى 3 أنواع وهي الإنحدار الخطي والذي يكون شكله خط مستقيم أما إذا كان شكله منحنيا فيسمى بالإنحدار متعدد الحواف، ويوجد أيضا الإنحدار اللوجستي (لحلح، 2020، صفحة 75)

ب- التعلم دون إشراف: "يحاول نظام التعلم إستنتاج هيكل البيانات غير المعلمة (ترك خوارزمية التعلم للإعتماد على نفسها في إستكشاف هيكل مدخلاتها وذلك لإكتشاف الأنماط الخفية في البيانات)" (عثمان ع.، 2018، صفحة 8)

أي ترك الآلة تتعامل بمفردها مع كومة من البيانات غير المصنفة وإكتشاف الأنماط والعلاقات بينها بحيث لا يوجد معلم أو مشرف يساعدها على تصنيف تلك البيانات لذا سيتوجب عليها ترتيبها وتصنيفها تلقائيا بالإعتماد على الترابطات والتشابهات الموجودة فيها وفي مايلي بعض أنواع التعلم غير الموجه أو التعلم دون إشراف:

- التجميع: تقوم الآلة بتقسيم البيانات حسب ميزات غير معروفة وتقوم بفرزها حسب ما تراه مناسبا فهي تحاول العثور على الكائنات المتشابهة حسب بعض الخصائص ودمجها في مجموعات والشكل التالي يوضح بإختصار طريقة التجميع:

الشكل 6.1: توضيح لطريقة التجميع



المصدر: (محمد لحلح،2020،ص77)

ومن بين أكثر الخوارزميات الشائعة في عملية التجميع هي:

K-mean-Clustering؛

Mean-Shift؛

.Dbscan

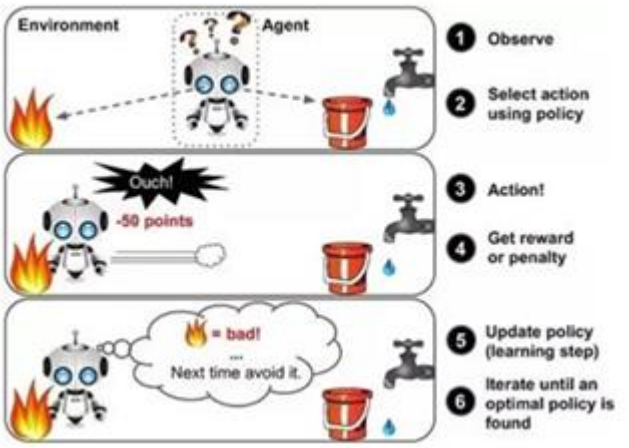
ومن بين أشهر تطبيقات طريقة التجميع هي تجميع الصور في ألبومات في تطبيق Google photos من خلال البحث عن الوجوه المميزة في الصور من خلال البحث على ميزات ومعرفة عدد الأفراد وإنشاء ألبومات وفقا لذلك (لحلح، 2020، صفحة 78/77)

- تعلم قواعد الربط: Association rule learning تستخدم هذه الطريقة لإكتشاف الأنماط في تدفق الطلبات مما يساعد على تحليل السلوكيات و العلاقات بين العناصر المختلفة، كما يمكن إستخدامها في تحليل عربات التسوق الإلكترونية لمعرفة المنتجات التي يتم شرائها معا مما يساعد على وضع خطط تسويقية ذات نجاعة وفعالية
- تقليل الأبعاد: Demontionality Reduction وتعرف أيضا بالتعميم ويقصد بها تجميع ميزات محددة ضمن ميزات أعم وأشمل منها يتم تطبيقها عادة في:
  - أنظمة التوصية؛
  - المحاكاة؛
  - تحليل الصور المزيفة؛
  - إدارة المخاطر.

تهدف هذه الطريقة إلى تقليل عدد المميزات أو features الموجودة في البيانات مع الحفاظ على أكبر قدر ممكن من المعلومات المهمة بهدف تحسين الآداء وتقليل التعقيد الحسابي؛

ب- تعلم الآلة المعزز: يظن البعض أن هناك نوعين من تعلم الآلة فقط وهما النوعين السابقين إلا أنه يوجد نوع ثالث يسمى بالتعلم المعزز أو التعلم التعزيزي وهذا النوع يقصد به أن الآلة تتعلم من خلال المكافآت والعقوبات مثل البشر تماما أي أنه عندما تقوم بتصرفات صحيحة تتم مكافئتها والعكس إن كان تصرفها خاطئا أو ناقصا تتم معاقبتها وذلك من خلال نقاط مثلا تزيد وتنقص حسب نوع تصرفها، كما يمكن تعريف التعلم المعزز بأنه "(تفاعل البرنامج مع بيئة ديناميكية. الهدف، هو تحقيق غاية معينة دون معلم ينبئه حتى بإقترابه من غايته تلك)" (عثمان ع.، 2018، صفحة 8) يمكن فهم التعلم المعزز بشكل أكثر من خلال الشكل التالي:

الشكل7.1: توضيح لطريقة التعلم المعزز



ا

المصدر: (عبد الرحمان عثمان، 2018، ص15)

يوضح الشكل السابق كيفية تعلم الوكيل من خلال التعامل مع البيئة وإتخاذ القرارت بناءا على المكافآت والعقوبات التي يحصل عليها

- الخطوة1 (observe): يقوم الوكيل Agent بمراقبة بيئته الخالية وجمع المعلومات عنها؛
- الخطوة 2(select action using policy): يختار الوكيل الإجراء الذي سيتخذه بناءا على سياسة التعلم التي يتبعها؛
- الخطوة 3(Action): يتم تنفيذ الإجراء المحدد؛
- الخطوة 4(Get Reward Or Penalty): تتم مكافآة الوكيل أو معاقبته حسب صحة تصرفه أو خطأه؛
- الخطوة5(update policy) يتم تحديث سياسة التعلم بهدف تحسين إتخاذ القرارات مستقبلا؛

-الخطوة 6 (Get reward or penalty): يستمر الوكيل في عملية التكرار حتى يتعلم الإستراتيجية المثلى التي تحقق له أفصل النتائج.

الروبوت في هذه الحالة في مواجهة بين فتح الماء والإقتراب من النار في حالة إقترابه من النار سيعاقب بإنقاص 50 نقطة مما يجعله يعرف أن هذا التصرف غير جيد وهذا ما يميز التعلم المعزز أنه يعتمد على التجربة والخطأ لتحسين الأداء.

المطلب *الثالث : طرق التنبؤ بإستخدام تعلم الآلة*

يعتبر التنبؤ والتخطيط للمستقبل أمرا هاما لدى جميع المؤسسات فهو يمكنها من إتخاذ القرارات الإستباقية، حيث توجد العديد من الطرق والأساليب والأدوات التي يمكن الإعتماد عليها لإجراء عملية التنبؤ كالطرق الإحصائية المعتادة أو الكلاسيكية مثل الإ نحدار بأنواعه، نماذج ARIMA والتي تعتمد أساسا على العلاقات الرياضية، إلا أنه و حتى تكون هذه الطرق ذات كفاءة أكثر فإنه يتوجب عليها الإلتزام بفرضيات معينة، و مع مرور الوقت تطورت أساليب التنبؤ وأصبحت تعتمد على مجال الذكاء الإصطناعي وتعلم الآلة بالإعتماد على خوارزميات متمكنة من التعامل مع البيانات دون التقيد حسب أي فرضيات وكما ذكرت سابقا أن هناك نوعين من الخوارزميات الأولى خاصة بالتصنيف وأخرى خاصة بالإنحدار، إلا أنه في هذا المطلب سنحاول التطرق إلى خوارزميات التصنيف فقط التي سيتم الإعتماد عليها في هذه الدراسة لبناء النموذج والمتمثلة في:

أولا: الإنحدار اللوجستي: يعتبر الإنحدار الوجستي أحد أكثر أنواع الإنحدار شيوعا في مجال تحليل البيانات فهو أسلوب وصفي يعتمد بالأساس على فهم العلاقة بين المتغير التابع والمتغيرات المستقلة الأخرى التي يمكن أن تكون متغير واحد أو عدة متغيرات مستقلة يمكن للإنحدار اللوجستي أن ينمذج متغير ثنائي الحد (Binomial Distribution) أي يأخذ المتغيرين قيمتين فقط أو إحتمالين فقط مثل (1/0) أو (بقاءالزبون / مغادرة الزبون) دالا بمجموعة من المتغيرات العشوائية المتوقعة سواء كانت هذه المتغيرات رقمية أو فئوية (خنان خضاري و مهدي محمود، 2003، صفحة 39)

يعتبر الإنحدار اللوجستي أسلوبا مهما للتنبؤ بسلوك الزبون من خلال البيانات المتوفرة يمكن معرفة ما إذا كان الزبون سيشتري المنتج أو لا إذ يمكن إستخدامه داخل برمجية Python مثلا لإنشاء نموذج تنبؤي.

ثانيا: شجرة القرار Decision Tree

تعتبر من بين أكثر خوارزميات التعلم الآلي شيوعا بسبب بساطتها وسهولة فهمها إذ يمكن إستخدامها كخطوات في تحليل القرارات وتمثيل وإتخاذه بإستخدام نموذج شبيه للشجرة بهدف الوصول إلى قرار نهائي، تعتبر شجرة القرار نموذجا مرنا يمكنها أن تتعامل بالبيانات الفئوية والرقمية أيضا. لقد كانت شجرة القرار عبارة على إقتراح من قبل خبير الإحصاء ليو بريمن في جامعة كاليفورنيا حيث تمثلت فكرته في عرض البيانات على شكل شجرة تكون فيها مجموعة من العقد الداخلية القائمة على شرط معين و تكون النتيجة في شكل فروع وكل ورقة تحمل تسمية فئة و تتميز شجرة القرار ب:

-تستطيع شجرة القرار مقاومة الأخطاء في حالة كانت بيانات التدريب تحتوي على أخطاء، كما أنه تستطيع معالجة هذه الأخطاء؛

-قواعدها مفهومة وبديهية، كما تتميز طريقة عملها بالسهولة والبساطة ويمكن لأي شخص فهمها؛

-تستطيع التعامل مع المتغيرات المستمرة و المتغيرات المتقطعة أيضا؛

-عدم الحاجة إلى توحيد البيانات أو تحجيمها؛

-لا تملك أي فرضيات مثل خطية البيانات مما يجعل إستخدامها ممكنا حتى في البيانات غير الخطية؛

-لا تملك أي حساسية تجاه البيانات المفقودة (وازن، تعلم الآلة وعلم البيانات).

تتكون شجرة القرار من:

1 – عقدة الجذر root node وهي أساس الشجرة؛

و تمثل نقطة البداية التي تتفرع منها البيانات الأخرى؛

2– العقد الداخلية Internal Node: تمثل نقاط إتخاذ القرار وتقع بين عقدة الجذر والعقد الطرفية؛

3–الفروع Branches : وهي الخطوط التي تربط بين العقد وتكون فيها احتمال وقوع العقد الداخلية.

ثالثا : الغابات العشوائية random forest

هي عبارة على مجموعة من أشجار القرار التي تعمل بشكل جماعي من أجل تحسين دقة النموذج وتقوم بعملية التنبؤ ثم تدمج نتائج القرار للحصول على نتيجة تنبؤ واحدة، تتميز خوارزمية الغابات العشوائية random forest بعدة مميزات مثل:

-إمكانية إستخدامها في التصنيف أو الإنحدار نظرا لسهولة تعاملها معهما؛

-تستخدم بشكل مباشر في المشاكل ذات الأبعاد العالية؛

-تعتبر سريعة نسبيا في التدرب و التنبؤ؛

-تعتمد على معلمة واحدة أو إثنتين للتوليف؛

-ترجيح الطبقة التفاظلية؛

-تضمين القيم المفقودة (Cutluer, 2011, p. 1).

يمكن شرح طريقة عمل الغابات العشوائية من خلال الخطوات التالية :

الخطوة الأولى : لنفترض أن لدينا مجموعة من البيانات في شكل جدول أي أن الجدول يحتوي على أعمدة و صفوف بحيث يمثل هذا الجدول البيانات الأولية أو البيانات الأصلية، كأول خطوة يجب إنشاء مجموعات فرعية (جدوال) بطريقة عشوائية من خلال أخذ عينات من البيانات الأصلية (الجدول الأول) مع إحتمال تكرار البيانات وذلك بهدف تنويع البيانات في كل شجرة بحيث يسمح من تقليل الخطأ وزيادة القدرة على التعميم وجعل النموذج التنبؤي أكثر دقة؛

الخطوة الثانية: بعد إنشاء الجداول الفرعية يتم إنشاء أشجار القرار علما أن كل شجرة تكون لها ميزات وسمات مختلفة عن الأخرى؛

الخطوة الثالثة : تقوم كل شجرة بعملية التنبؤ بطريقة منفردة بالإعتماد على البيانات المسندة لها والتي تم تدريبها عليها ونظرا لأن بيانات كل شجرة تختلف عن الأخرى فهذا يعني أن نتائج التنبؤ ستكون مختلفة عن بعضها إما بصفة جزئية او كلية؛

الخطوة الرابعة : تعتبر هذه الخطوة أساس القرار بحيث يتم دمج نتائج التنبؤ السابقة وإختيار نتيجة التنبؤ من خلال التصويت بالنتيجة الغالبة والتي تكررت أكثر.

رابعا: خوارزمية التدرج التعزيزي **Gradient Boosting Machine – GBM**

ويمكن تسميتها أيضا بخوارزمية التدرج المعزز وهي تقنية تعلم آلة ضمن فئة خوارزميات التعزيز التي تبني نموذجًا تنبؤيًا قويًا عن طريق تجميع نماذج تعلم ضعيفة (عادةً أشجار قرار قصيرة). يتم تدريب هذه النماذج بشكل تكراري، حيث يركز كل نموذج لاحق على تقليل الأخطاء المتبقية (residuals) التي ارتكبها النموذج السابق. تستخدم خوارزمية هبوط التدرج لتحسين النموذج في كل خطوة عن طريق تعديل تنبؤات النموذج الضعيف بما يتناسب مع تدرج دالة الخسارة. هذا النهج يسمح للنموذج النهائي بالتقاط علاقات معقدة في البيانات وتحقيق دقة تنبؤية عالية. (Trevor Hastie و Robert Tibshirani، 2008، صفحة 395)

المبحث الثالث: الدراسات السابقة

المطلب الأول: الدراسات العربية

الدراسة الأولى

حنان خضاري (2023) " المقارنة بين تعلم الآلة والتعلم العميق في بناء نموذج قادر على التنبؤ بخروج العملاء من الشركات الخدمية"

هذه الدراسة عبارة على مقال مفتوح تم نشره في المجلة المصرية للسكان وتنظيم الأسرة في سنة 2023، حيث هدفت هذه الدراسة إلى إنشاء نموذج يمكن من خلاله التنبؤ بسلوك الزبائن المتمثل في خروجهم من الشركات الخدمية من خلال المقارنة بين طرق التعلم الآلي والتعلم العميق بإعتبار أن هذه التقنيات الحديثة تمنح ميزات جديدة ومتطورة مقارنة بالطرق التقليدية للتنبؤ تم تطبيق المنهج الكمي التطبيقي في هده الدراسة من خلال تطبيق عدة خوارزميات على بيانات من شركة خدمية مثل شجرة القرارDecision Tree وتعزيز التدرج ونماذج الإنحدار والجار الأقرب KNN والغابات العشوائية Random Forestوالشبكات العصبية على مجموعة من البيانات التي تم تحميلها من موقع kaggle بغرض البحث العلمي وبإستخدام على برنامج orange تم تحليل هده البيانات والمقارنة بين الخواررزميات السابقة بناءا على المعايير التالية(Recall، Classification Accuracy، Precision، Area Under the Curve) و تم التوصل إلى أن طريقة التدرج المعزز هي الطريقة الأفضل التي تخدم هذه الدراسة بإنشاء نموذج تنبؤي ذو دقة بنسبة 84.7% كما أوصت هذه الدراسة بإستخدام طرق أكثر دقة وذلك من أجل إتخاذ القرارت المناسبة . (محمود، 2023)

الدراسة الثانية

حسين مصيلحي سيد_ إيناس محمد العباسي (2023) " دور تحليلات البيانات الضخمة في التنبؤ بنية تحول العملاء بإستخدام تحليل المشاعر "

هذه الدراسة هي مقال مفتوح نشر في مجلة الدراسات التجارية المعاصرة العدد السابع عشر الجزء الثاني يناير سنة 2023 ، حيث سعت هذه الدراسة الى معرفة دور تحليلات البيانات الضخمة Big Data في التنبؤ بنية تغيير العملاء لعلامات أخرى منافسة أي تغيير مزود خدماتهم خلال فترة معينة وذلك من خلال إجراء دراسة تطبيقية في عدة شركات للهاتف المحمول في مصر من خلال إجراء حملات إعلانية على منصة التواصل فيسبوك وجهت لمستخدمي الهاتف المحمول تم جمع 384 مفردة من المجتمع ككل والمتمثل في الشركات (فودافون، أورانج، إتصالات، WE) وبما أن إمكانية الشركة التنبؤ بفقدان عملائها أومعرفة نيتهم بالمغادرة يعتبر مكسبا لها كونها يمكنها التحسين من تجربة عملائها للمحافظة عليهم أكثر، ويعتبر تحليل البيانات الضخمة Big Data Analytics وسيلة جد فعالة في مثل هذه الدراسات، ركز الباحث في

هذه الدراسة على تحديد أبعاد البيانات الضخمة التي تساعد في إجراء عملية التنبؤ وكذلك إستكشاف الفرص المتاحة التي يمكن أن نطبق فيها أسلوب تحليل المشاعر لمسخذمي منصات التواصل الإجتماعي وإستخدامها في التنبؤ بسلوك العملاء و ولائهم لعلامة تجارية دون أخرى وذلك من خلال تحليل الأصوات ومقاطع الفديو والنصوص ، إعتمد الباحث في هذه الدراسة على منهجين مختلفين؛ المنهج الوصفي بهدف التعرف على مصطلحات الدراسة والمنهج التفسيري أو السببي لمعرفة العلاقة بين المتغيرات المستقلة والتابعة وذلك بإستخدام عدة أدوات إحصائية مثل(ألفا كرونباخ Cronbach's Alpha ، التكرارات Frequencies، النسب المئوية Percent، الوسيط الحسابي Mean) وتوصلت الدراسة إلى أن تحليل البيانات الضخمة يساعد على إتخاذ قرارات أفضل من خلال الإعتماد على بيانات الزبائن وتحليل مشاعرهم مثل رضاهم حول منتج معين يؤدي الى زيادة فرصة الإحتفاظ بهم، كما يمكن تحويل الزبائن المحتملين إلى زبائن مشترين. (النجار و حسين مصيلحي سيد، 2024)

الدراسة الثالثة

نهى سامي، إبراهيم عامر " تأثير إستخدام أدوات الذكاء الإصطناعي في التنبؤ بسلوك المستهلك " دراسة شبه تجريبية في إطار نموذج قبول التكنولوجيا (2023)

هذه الدراسة هي مقال مفتوح نشر في مجلة البحوث الإعلامية 70 المجلد 2 حيث تهدف هذه الدراسة إلى معرفة تأثير مختلف أدوات الذكاء الإصطناعي في التنبؤ بسلوك المستهلك ودوره الهام في تعزيز وتحسين العلاقات بين الشركة والزبائن ، وكذلك التعرف على كيفية إستخدامه في التنبؤ ومدى تأثيره على سلوك الزبون وإختبار مدى نجاعة و فعالية الأدوات الخاصة بالذكاء الإصطناعي؛

إستعملت هذه الدراسة المنهج الشبه التجريبي على مجتمع دراسة متمثل في طلاب كلية الإعلام وتكنولوجيا الإتصال في جامعة مصر للعلوم والتكنولوجيا بإستخدام الإستبيان وتم إختيار عينة متكونة من 40 مفردة، ويعود سبب إختيار العينة لعدة عوامل من بينها أن الفئة العمرية المتمثلة في الشباب هي أكثر فئة تعاملا مع التكنولوجيا ولها رغبة دائمة في معرفة وتعلم المزيد في هذا المجال، وهذا ما يساعد على فهمهم للموضوع بنسبة أكبر من باقي الفئات العمرية الأخرى، تم إستعمال عدة أدوات خاصة باذكاء الإصطناعي وهي (Phrase.com/ai, Xineoh.com, Chatfuel.com/ai, Acrolinx.com, Semrush.com) تم تخليل نتائج الدراسة بإستخدام برنامج تحليل الحزم الإحصائية Spss حيث توصلت هذه الدراسة إلى أن كل أداة من هذه الأدوات تميزت في إختصاص مختلف عن الأخرى وذلك بالإعتماد على إختبار الأدوات من حيث ماهية كل أداة على حدى، أهدافها، الخدمات التي تقدمها الأداة، القنوات والمنصات التي تعتمد عليها، عملاء كل

أداة ومن يستخدمها ، طريقة إستعمالها إن كانت مدفوعة أو مجانية، وغيرها من المعايير التي تم الإعتماد عليها في هذه الدراسة. (عامر، 2024)

المطلب الثاني: الدراسات الأجنبية

الدراسة الأولى

Manish Kumar Sharma, et all (2023).

هذه الدراسة عبارة على مقال مفتوح نشر في مجلة International Research Journal of Engineering & Applied Sciences (IRJEAS) تمحورت هذه الدراسة حول التنبؤ بفقدان الزبائن بإستخدام تقنيات تعلم الآلة القائم على تحسين إختيار الميزات وأكدت على أهمية الطريقة المستخدمة في إختيار البيانات المتعلقة بالدراسة ودور التعلم الآلي في دعم إتخاذ القرار خاصة في مجال الأعمال إستخدمت هذه الدراسة المنهج التحليلي التجريبي من خلال تطبيق عدة خوارزميات خاصة بالتعلم الآلي من بين هذه الخوارزميات أشجار القرار الإنحدار اللوجستي و الغابات العشوائية والمتجهات الداعمة وذلك من خلال الإعتماد على مجموعة من البيانات التي تخص العملاء مثل تاريخ تعاملهم مع الشركة وكذلك ملصقات الإنفصال churn labels حيث قامت هذه الدراسة بتطوير نموذج قادر على التنبؤ بسلوك الزبائن المحتملين وغير المحتملين بدقة أكثر تميزت هذه الدراسة بكونها قامت بإستخذام تقنية لإستبعاد التكرار أو مايعرف ب (Recursive Feature Elimination - RFE) بهدف تحسين دقة النموذج وكذلك محاولة تقليل التعقيد الحسابي كما تم تطبيق 3 نماذج تنبؤية في هذه الدراسة وقد حققت جميعها نسب دقة فوق 79 % وتمت المقارنة بينهم وكان نموذج KNN حسب معظم معايير التقييم هو الأفضل من بينهم وتم التوصل إلى أن إستخدام هذه التقنية يساعد في تحسين دقة النموذج الخاص بالتنبؤ أي أنها طريقة جد فعالة، كما أوصت هذه الدراسة إلى إستخدام KNN بإعتباره أفصل نموذج للتنبؤ بمغادرة الزبائن بنسبة دقة 81 Sharma) %. & Vaibhav Patel, 2023)

الدراسة الثانية :

Predicting online user behaviour using deep learning algorithms"Armando Vieira (2016)"

هذه الدراسةعبارة على ورقة بحثية نشرت على منصة arXiv في 27/05/2016 سعت هذه الدراسة إلى التنبؤ بالسلوك الشرائي لدى المستخدمين لمنصات التجارة الإلكترونية وذلك بإستخدام تقنيات التعلم العميق Deep Learning، حيث تم جمع البيانات المتعلقة بهذه الدراسة من السجلات الخاصة

بمستخدمي منصات التجارة الإلكترونية والتي توضح تفاعلهم لمدة ستتة أشهر، تتضمن البيانات متغيرات مختلفة مثل: (عدد النقر على الإعلانات، عرض المنتج، عرض سلة الشراء) جمعت هذه البيانات بشكل يومي من قبل 7469 مستخدم وركزت الدراسة على التنبؤ بالسلوك الشرائي للمستخدمين بإستهدافها لتحليل سلوك المستخدمين في بيئة التجارة الإلكترونية مع مراعاة أنماطهم المختلفة إعتمدت الدراسة على المنهج الوصفي و التجريبي بإستخدام أدوات التعلم العميق المتمثلة في (الشبكات العصبية العميقة Deep Neural Networksو شبكات الإعتقاد العميق Deep Belief Networks)، تميزت هذه الدراسة بعدم توازن بياناتها وذلك راجع إلى تنوعها لهذا تم إستخدام تقنية word2ves لمعالجة البيانات، توصلت الدراسة إلى أن أدوات التعلم العميق تفوقت على باقي الأساليب التقليدية الأخرى في التنبؤ بسلوك الزبائن خاصة في مجال التجارة الإلكترونية حيث أن خوارزمية شبكات الإعتقاد العميق إستطاعت التنبؤ بدقة بلغت 84 %كما بلغت دقة التنبؤ حسب نموذج الشبكات العصبية العميقة بنسبة 86% (Redzebra, 2016) .

الدراسة الثالثة:

Varoon Raja,et all
"Predicting Customer Behavior in E-Commerce Using Machine Learning Algorithms: A Mathematical Approach "

هذه الدراسةهي عبارة على ورقة بحثية تم نشرها في منصة Easy Chair Preprints تهدف هذه الدراسة إلى بناء نموذج يمكن من خلاله التنبؤ بالسلوك الشرائي لدى الزبائن بإستخدام عدة نماذج تعلم آلي بهدف فهم إتجاهات الزبائن بشكل أفضل وأيضا تحسين إستراتيجيات التسويق، حيث تم الإعتماد في هذه الدراسة على بينات متنوعة منها ديموغرافية مثل العمر ، الجنس، وبيانات إقتصادية مثل الدخل السنوي ومنها بيانات سلوكية مثل عدد المشتريات، فئة المنتج والوقت الذي يستغرقه الزبون في التصفح،...إلخ، كما تم معالجة البيانات المتعلقة بالدراسة مسبقا بعدها تم تحليل البيانات وإختيار الميزات (المتغيرات) المناسبة لبناء النموذج بعدها تدريبه وتقييم ستة نماذج تعلم آلي مختلفة تتمثل في الإنحدار اللوجستي، التعزيز المتدرج، الغابات العشوائية، آلة دعم المتجهات، أقرب جار، تتم عملية تقييم النماذج حسب معايير خاصة بالتأكد من دقة النموذج مثل معيار F1 Scor و Precision و Recall وتوصلت هذه الدراسة إلى أن نماذج التجميع فعالة للغاية في حالة التنبؤ بسلوك الزبائن خاصة خوارزمية الغابات العشوائية وخوارزمية التدرج المعزز حيث تفوقوا على باقي النماذج من حيث الدقة، يوفر التنبؤ بالسلوك الشرائي لدى الزبائن رؤية واضحة للشركات لتخصيص جهودها التسويقية بشكل أفضل وإستهداف الزبون بشكل مباشر كما ترى

هذه الدراسة أنه يمكن للأبحاث المستقبلية إستكشاف سمات إضافية ونماذج أكثر تطورا قادرة على التنبؤ. (raja & Jumes Kung, 2025)

الدراسة الرابعة: Desislava koleva"Machine learning methods applicable in customer lifecycle management "2024

مقال مفتوح تم نشر في مجلة Izvestia Journal Of The Union Of Scientists _VARNA المجلد 13 العدد 1 تناولت هذه الدراسة إلى بناء نموذج تعلم إلي يهدف الى التنبؤ بالسلوك الإستهلاكي للزبائن من خلال مدة تعاملهم مع الشركات تم إستخدام المنهج الوصفي التحليلي وهدف نموذج هذه الدراسة إلى توقع سلوك الزبائن من خلال تصنيف سلوكهم إلى فئتين إما تفاعلهم مع الحملات التسويقية أو مغادرتهم، وذلك بالإعتماد على طرق تعلم الآلة الخاضعة للإشراف مثل شجرة القرار، الدعم الناقل، الإنحدار اللوجستي، الشبكات العصبونية، وغيرها من خوارزميات التعلم وإعتمدت هذه الدراسة على مؤشرين أساسين في تقييم النموذج وهما مؤشر الدقة العامة Accuracy و ومرشر Auc ومن خلال المقارنة بين هذه النماذج حيث أضهرت نتائج المقارنة أن أكثر دقة قد توصل إليها النموذج كانت % 94.6 بإستخدام خوارزمية AdaBoostبينما حققوا باقي النماذج نتائج قوية أيضا خاصة الإنحدار اللوجستي والشبكات العصبونية دقة تجاوزت 91% إضافة إلى أن الدراسة أشارت إلى أهمية تحسين أداء النماذج وذلك من خلال تعديل البيانات غير المتوازنة وظبطها. (koleva، 2024)

الدراسة الخامسة:

Sumnana Sharma Poudel, Suresh Pokharel, Mohan Timilsina "Explaining Customer Churn Prediction in Telecom Industry Using Tabular Machin Learning Models" 2024

هذه الدراسة عي عبارة على مقال مفتوح نشر في مجلة Machin Learning with applications وهي مجلة من الصنف A العدد 17 هدفت هذه الدراسة إلى التنبؤ بمغادرة الزبائن من شركة الإتصالات وذلك بإستخدام قاعدة بيانات **telecom_customer_churn.csv** خاصة بشركات الإتصال من موقع Kaggleمن خلال بناء نموذج تنبؤ بإستخدام تقنيات تعلم الآلة، حيث إعتمدت على المنهج الوصفي التحليلي، وذلك بإستخدام عدة خوارزميات خاصة بالتصنيف تتمثل في (خوارزمية التدرج المعزز GBM،شجرة القرار، الإنحدار اللوجستي، الغابات العشوائية،SVM، الشبكات العصبية، Ada Boost،XG Boost) وتوصلت هذه الدراسة إلى أن أفضل نموذج من بين النماذج التي تم تجريبها في عملية التنبؤ كان نموذج

GBM حيث وصلت نسبة دقة النموذج Accuracy في التنبؤ بمغادرة الزبائن ل81 %. (Shurma & Suresh Pokharel, 2024)

خلاصة الفصل الأول:

في هذا الفصل تم التعرف على أهم الأسس النظرية المرتبطة بالسلوك الإستهلاكي للزبون والذي يعتبر التنبؤ به أداة جوهرية لا بد أن تملكها كل مؤسسة إقتصادية حيث تساعدها بشكل كبير على تحقيق هدفها الأساسي إضافة إلى التطرق إلى معرفة الخلفية النظرية لتعلم الآلة بصفة عامة، وأهم الخوارزميات التي يمكن من خلالها بناءنموذج تنبؤ يمكن للمؤسسات الإقتصادية الإعتماد عليه لتحسين إستراتيجياتها والعمل على زيادة عوائدها بشكل أكبر وأسرع، كما تم ذكر بعض الدراسات السابقة العربية والأجنبية و التي تمحورت حول نفس الموضوع وهو التنبؤ بسلوك الزبائن بإستخدام تقنيات تعلم الآلة إذ تجدر الإشارة إلى أن الدراسة الأجنبية الخاصة ب Sumnana Sharma Poudel, Suresh Pokharel, Mohan Timilsina التي هي بعنوان Explaining Customer Churn Prediction in Telecom Industry Using Tabular Machin Learning Models قد إستعملت نفس قاعدة البيانات التي قمنا بإستخدامها في هذه الدراسة ولذلك سنجري عملية مقارنة بين النتائج التي تحصل عليها صاحب الدراسة والنتائج التي ستحصل عليها بعد تطبيق النماذج التي سنعتمد عليهافي الفصل التطبيقي.

الفصل الثاني : الفصل التطبيقي

بعد ماتطرقنا في الفصل الأول من هذه الدراسة إلى الجوانب النظرية المتعلقة بالسلوك الإستهلاكي للزبون وأهم المحددات التي تدفعه لإتخاذ قراراته تجاه المنتجات والخدمات، سنتحدث في الفصل الثاني والمتمثل في الجانب التطبيقي للدراسة والذي سنقوم فيه بإنشاء نموذج تنبؤي قادر على توقع سلوك الزبون والمتمثل في تخلي الزبون عن الخدمة وإلغاء إشتراكه من شركة الإتصالات أو بقائه وذلك من خلال تحديد أهم العوامل والمتغيرات التي تؤثر في سلوك الزبون عن طريق تحديد الخصائص التي تؤثر بشكل أكبر في بناء النموذج والإعتماد عليها وذلك بإستخدام خوارزميات خاصة بالتعلم الآلي الخاضعة للإشراف، وبما أن البيانات المتاحة متنوعة مما يجعلها تتيح لنا فهم سلوك الزبون بشكل أفضل وأعمق، من خلال هذا الفصل سنقوم بعرض الخطوات المنهجية التي سنتبعها لإنشاء النموذج بداية من التعريف بالبيانات و مصدر جمعها والأدوات المستخدمة في الدراسة والمراحل التي سنتبعها لإنشاء النموذج إلى بناء النموذج وتقييمه وإجراء مقارنة بين هذه الدراسة ودراسة أخرى قامت بإنشاء نموذج للتنبؤ بإستخدام نفس البيانات.

المبحث الأول: إستكشاف بيانات الدراسة

المطلب الأول: التعريف بمتغيرات الدراسة ومصدر جمع البيانات

أولا: مصدر البيانات

تم الإعتماد في هذه الدراسة على قاعدة بيانات خاصة بشركة من شركات الإتصالات telecom_customer_churn.csv تم تحميلها من موقع Kaggle وهو موقع خاص بنشر البيانات التابعة للشركات والمؤسسات أو الجامعات التي تقوم بنشر بياناتها بهدف البحث العلمي ومساعدة الباحثين في وضع مقترحات أو حلول لمشكلة معينة.

رابط قاعدة البيانات : The Customer Churn dataset contains 7,043 customer data from a Telecommunications company **in California in 2022**.

ثانيا: وصف البيانات

تتكون البيانات من 36 متغير تشمل هذه المتغيرات 21 متغيرنوعي و15 متغير كمي، ويمثل آخر متغير Churn المتغير التابع Y أو ما يسمى ب Target Variable، أما باقي المتغيرات فهي متغيرات مستقلة و الجدول التالي يشرح المتغيرات التي تم الإعتماد عليها في الدراسة:

جدول1.2: وصف متغيرات الدراسة

المتغير	نوع المتغير	التفسير
Gender	نوعي	جنس الزبون
Age	كمي	عمر الزبون
Married	نوعي	الحالة الإجتماعية للزبون متزوج / أو لا
Number of Dependents	كمي	عدد الأشخاص الذين يعيلهم الزبون
City	نوعي	مدينة الزبون
Zip Code	كمي	الرمز البريدي
Latitude	كمي	خط العرض
Longitude	كمي	خط الطول
Number of Referrals	كمي	عدد الإحالات أو عدد التوصيات التي قام بها الزبون تجاه المنتج
Tenure in Months	كمي	مدة إشتراك الزبون بالأشهر
Offer	نوعي	العرض الذي حصل عليه الزبون مثلا خصم 25% أو باقة
Phone Service	نوعي	خدمات الهاتف
Avg Monthly Long Distance Charges	كمي	متوسط الرسوم الشهرية للمكالمات الدولية
Multiple Lines	نوعي	إمتلاك الزبون لأكثر من خط هاتف أي يملك عدة شرائح
Internet Service	نوعي	خدمة الإنترنت
Internet Type	نوعي	نوع الإنترنت الذي يستخدمه الزبون (شبكة هاتف4g أو وايفاي أو قمر صناعي ....)
Avg Monthly GB Download	كمي	متوسط تحميل البيانات شهريا بالجيغابايت
Online Security	نوعي	الأمان عبر الإنترنت (الحماية من الفيروسات، التشفير، المصادقة الثنائية،...)
Online Backup	نوعي	النسخ الإحتياطي لحماية البيانات

Device Protection Plan	نوعي	خط لحماية الأجهزة الإلكترونية من المشاكل التقنية أو المادية
Premium Tech Support	نوعي	الدعم الفني المميز (تكون الخدمات فيه أكثر تطورا)
Streaming TV	نوعي	إستعمال الإنترنت في مشاهدة البرامج أو محتوى إعلامي من خلال التلفاز
Streaming Movies	نوعي	إستعمال الإنترنت في مشاهدة الأفلام
Streaming Music	نوعي	إستعمال الإنترنت في الإستماع إلى الموسيقى
Unlimited Data	نوعي	البيانات غير المحدودة متاحة للزبون دون أي قيود
Contract	نوعي	العقد(يكون بين الشركة والزبون يتضمن شروط إلغاء الخدمة، مدة الإلتزام،..)
Paperless Billing	نوعي	الفوترة الرقمية (تكون الفاتورة إلكترونية بدون اوراق)
Payment Method	نوعي	طريقة الدفع (نقدا، خصم تلقائي ، بطاقة إئتمان ،.....)
Monthly Charge	كمي	الرسوم الشهرية (يدفعها الزبون شهريا)
Total Charges	كمي	الرسوم الإجمالية (الرسوم الشهرية +الظرائب والرسوم الحكومية)
Total Refunds	كمي	إجمالي المبالغ المستردة(المبالغ التي تعيدها الشركة للزبون بسبب إلغاءه للخدمة تم دفعها مسبقا أو أخطاء في الفوترة)
Total Extra Data Charges	كمي	إجمالي الرسوم الإضافية (الرسوم التي يدفعها الزبون في حالة لم تكفي البيانات الشهرية وقام بتجديد الحدمة في نفس الشهر)
Total Long Distance Charges	كمي	إجمالي رسوم المكالمات الخارجية
Total Revenue	كمي	مجموع العوائد التي حصلت عليه الشركة من الزبون
Customer Status	نوعي	حالة العميل (علاقته بالشركة وطلبه للخدمة دائم التعامل معها أو زبون معلق ولديه مشاكل في الدفع أو زبون مفصول أو زبون جديد فام الإنضمام حديثا)
Churn	نوعي	المتغير الهدف :يشير إلى ما إذا تخلى الزبون عن الخدمة أو لا

المصدر: من إعداد الطالبة بالإعتماد على مخرجات لغة Python

المتغير التابع: Churn يمثل حالة العميل تجاه الشركة هل سيغادر أم لا

المتغيرات المستقلة: كل المتغيرات الباقية هي متغيرات مستقلة

المطلب الثاني: الأدوات وبيئة التطوير المستخدمة في الدراسة

لقد تم الإعتماد في هذه الدراسة على عدة أدوات خاصة بتعلم الآلة وهي:

أولا: برمجية بايثون Python Program: هي لغة برمجة ذات إستخدام واسع النطاق في الوقت الحالي أنشأها المبرمج الهولندي (Guido Van Rossum) و قام بإطلاقها كأول مرة في سنة 1991، وتعني كلمة Python أحد أخطر أنواع الأفاعي

تتميز لغة Python بأنها تعتبر من بين اللغات القوية التي تتسم بسهولة قراءة تعليماتها وبساطتها مقارنة باللغات الأخرى مما يسمح للمبرمجين بسرعة تطوير البرامج، كما أنها قابلة للتنفيذ على أشهر برامج التشغيل مثل Mac و windows "تعتبر لغة Python لغة مثالية للمبتدئين وذلك لبساطتها حيث أنها تتطلب عدد اقل من الأسطر البرمجية لتنفيذ نفس المهمة مقارنة باللغات البرمجية الأخرى " (Taher, n.d, p. 01) ، تتعامل مع العديد من الملفات مثل ملفات Excel وملفات Csv وغيرها، و تحتوي برمجية Python على العديد من المكتبات التي تساعد في إنجاز العديد من المهام وفي أكثر من مجال مثل الرياضيات، تحليل البيانات، صناعة الألعاب، معالجة الصور و مقاطع الفديو إلا أننا سنتحدث في هذه الدراسة عن المكتبات الي سنعتمد عليها في إنشاء النموذج التنبؤي والمتمثلة في :

أ- مكتبة Numpy: هي مكتبة في Python خاصة بالتعامل مع المصفوفات وهي مكتبة تعتمد عليها معظم مكتبات التعلم الآلي والتعلم العميق كأساس لها، وتكمن أهمية مكتبة Numpy أنها تعتبر أسرع بكثير من باقي مكاتب Python إضافة إلى أنها تتميز بالدقة العالية وذلك يرجع إلى الطريقة التي تقوم فيها المكتبة بتخزين البيانات والتي تعرف ب Locality Of Reference، فهي تجمع العناصر من نفس النوع في مكان متصل بالذاكرة (Contiguous Memory Block) تتميز مكتبة Numpy بتقديم كائنات مصفوفة nd array، توفر وظائف رياضية متنوعة مثل الجبر الخطي والعمليا الإحصائية كما أنها تعالج البيانات المعقدة بكفاءة و بطريقة سهلة وفعالة، يمكن إستيراد المكتبة في لغة Python عبر الأمر التالي : import numpy as np

مكتبة pandas : تعتبر من بين المكتبات الشهيرة في Python وهي أداة أساسية في تحليل البيانات، تتعامل مكتبة pandas مع هيكل البيانات Series وهي عبارة عن مصفوفة أحادية الأبعاد وتتعامل أيضا مع Data frame وهي عبارة على جداول ثنائية الأبعاد تشبه جداول برنامج Excel، تتميز مكتبة

pandas بالمرونة مما يسمح لها بالتعامل مع أنواع مختلفة من البيانات مثل(الأعداد الصحيحة، الأعداد العشرية، التواريخ، السلاسل النصية،....) كما أنها تدعم قراءة البيانات من ملفات مختلفة مثل (Csv، Excel، Sql، Json) توفر مكتبة pandas أدوات متقدمة تساعد على تصفية البيانات وتنظيفها، تجميعها ودمجها ويمكن إستيراد مكتبة pandas في برمجية python بإستخدام الأمر التالي:

import pandas as pd

ب- مكتبة Scikt Learn: هي مكتبة من أهم مكتبات لغة Python تستعمل مكتبة Scikt Learn للتصنيف والإنحدار والتجميع وتقوم بتقسيم البيانات وتدريبها وإختبارها وأيضا إختيار النموذج وتحسينه وتملك أدوات تقييم الأداء كما تعمل على معالجة البيانات من خلال التعامل مع البيانات المفقودة وحساب مصفوفة التحقق المتقاطع Confusion matrix كما أنها تعتمد كثيرا على المكتبات السابقتين وتتميز بتوفيرها لنماذج الإنحدار المختلفة مثل شجرة القرار، الإنحدار اللوجستي، K-Means وسهولة إستخدامها إضافة إللى أنها توفر بيانات جاهزة يمكن التجريب عليها ويمكن إستيراد المكتبة بإستخدام الأمر التالي: import Scikt_Learn as

ت- مكتبة matplotlib: هي مكتبة تستخدم لرسم الرسومات البيانية في Python تسمح برسم مخططات ثنائية الأبعاد مثل (Line Plot)،(Bar Charts)،(Pie Charts)،(Scatter Plots) تتمثل أهم خصائصها في المرونة العالية لأنها تستطيع تخصيص كل جزء من الرسم مثل اللون والخطوط والعناوين وغيرها، كما أنها تدعم الصور وتتحكم بكل الرسومات أي يمكن إنشاء عدة رسومات في نافذة واحدة وتصديرها بصيغ مختلفة PDF، SVG، PNG ويتم إستيراد المكتبة بإستخدام الأمر التالي

: import matplotlib as plt

ث- مكتبة Seaborn: هي مكتبة تقوم بتصوير البيانات بلغة Python مبنية أساسا على مكتبة matplotlib توفر هذه المكتبة واجهة ذات مستوى عالي لتصميم الرسومات الإحصائية بطريقة جذابة وغنية بالمعلومات(Seaborn, 2025)

ثانيا: بيئة التطوير

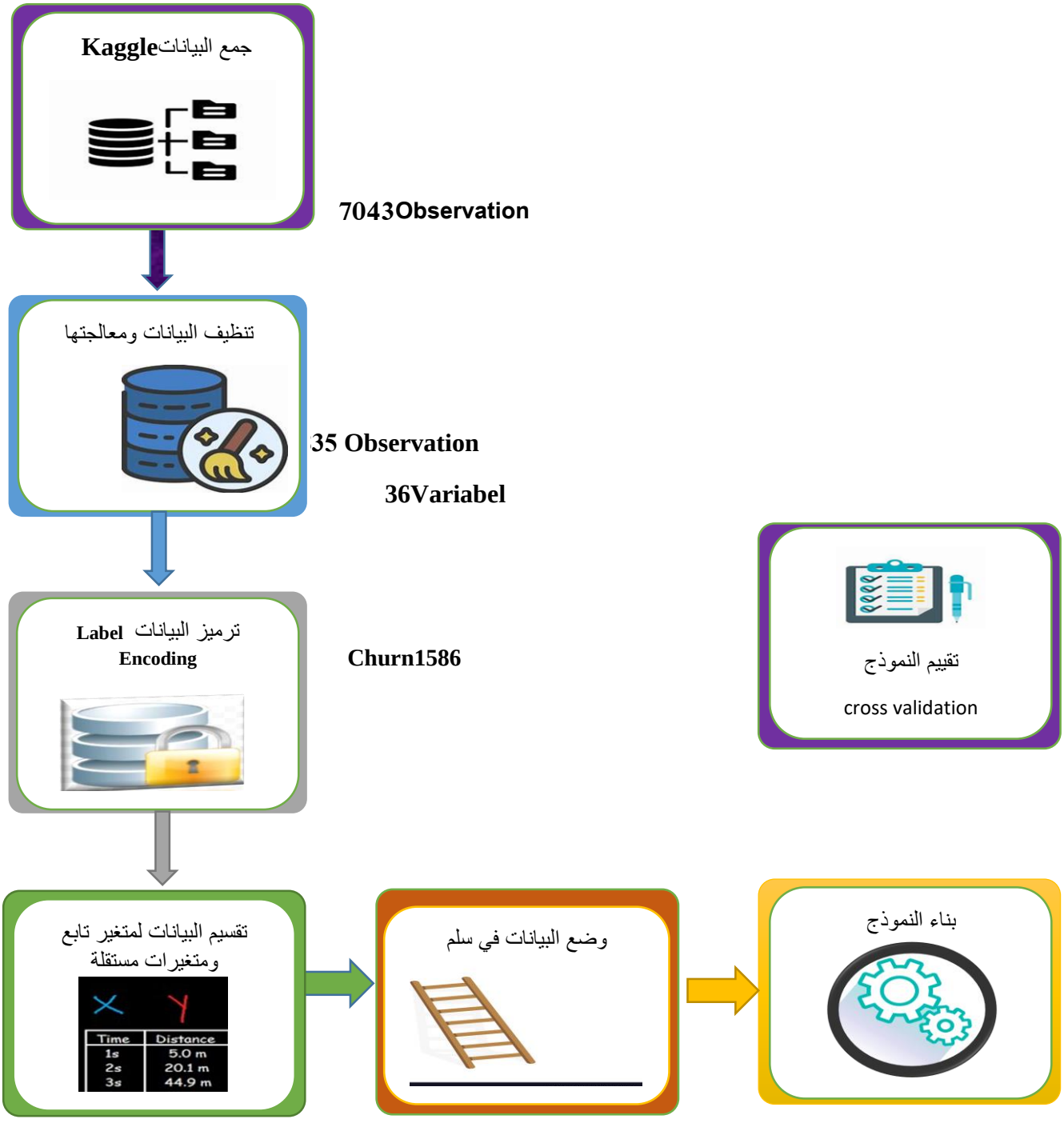
Jupyter Notebook: هي أداة قوية جدا في التعامل مع المشاريع المتعلقة بالتعلم الآلي التي تحتوي على النصوص ومقاطع الفديو، الصور، وتطويرها كما أنها تسمح بتنفيذ العمليات البرمجية والعمليات الرياضية والقيام بالرسومات التوضيحية والنصوص السردية وجمعها في مستند واحد (وازن، التعلم العميق من الأساسيات ، حتى بناء شبكة عصبية عميقة بلغة البايثون ، 2022، صفحة 47) كما

يمكن أن نعرف بيئة التطوير Jupyter Notebook على أنها برنامج يستخدم بشكل واسع في تعلم الآلة والبرمجة و يتميز بكونه يدعم عدة لغات برمجة مثل لغة Python ولغة R كما أنه يتناسب بشكل كبير مع أدوات تحليل البيانات مما يمنحه سهولة في معالجة البيانات إضافة إلى ذلك فإنه يملك القدرة على التعامل مع المكتبات الخاصة بلغة Python أي يمكن إجراء عملية التنبؤ.

المطلب الثالث: خطوات بناء النموذج وأدوات تقييمه

بعد التعرف على البيانات الخاصة بالدراسة ومصدر البيانات في المطلبين السابقين وكذلك معرفة الأدوات المستخدمة وبيئة التطوير التي سيتم الإعتماد عليها في إنشاء النموذج التنبؤي، لا بد من معرفة الخطوات أو المراحل الخاصة بهذا النموذج والتي تعتبر كمنهجية لشرح كيفية عمل النموذج بدءا بجمع البيانات وإختيار أنسب المتغيرات إلى غاية تقييم نموذج تعلم الآلة، و يساعد المخطط التالي على توضيح آلية عمل بناء النموذج الخاص بهذه الدراسة:

الشكل8.2: خطوات بناء النموذج



المصدر: من إعدادا الطالبة

أولا: شرح خطوات بناء النموذج

_ جمع البياناتData Colction : و يقصد بجمع البيانات الحصول على البيانات الخام التي سنستخدمها في الدراسة لبناء نموذج للتنبؤ بمغادرة الزبائن، حيث تعتبر البيانات أساس كل دراسة ويمكن جمعها بعدة طرق ومن مصادر مختلفة مثل ملفات (Excel/ Csv) أو يدويا وفي هذه الدراسة تم جمعها من مصدر مفتوح وهو موقع (Kaggle) ؛

_ تنظيف البيانات ومعالجتها Data Cleaning and Preprocessing : في هذه المرحلة يجب تهيئة البيانات وتحضيرها لبناء النموذج وهي خطوة جد مهمة بعد عملية جمع البيانات حيث يجب تنظيف البيانات من القيم الشاذة و القيم المتكررة و القيم المفقودة ومعالجتها إما بالحذف أو التعويض والتأكد من صلاحية قاعدة البيانات؛

_ ترميز البيانات Data Encoding: بما أن خوارزميات التعلم الآلي لا تستطيع أن تتعامل مع البيانات النصية وجب ترميز البيانات أي تحويل البيانات النصية أو الفئوية إلى بيانات رقمية؛

_ وضع البيانات في سلم Data Scaling : أي وضع البيانات في مجال موحد مثال العمر[2، 2−] لأن نماذج التعلم الآلي تكون دقيقة اكثر عندما تكون البيانات موحدة وفي نفس النطاق أو المجال؛

_ تحديد المتغيرات (السمات) Features Selection : يقصد بها تحديد المتغيرات المستقلة الأكثر إرتباطا بالمتغير التابع والتي تساهم في إعطاء أدق النتائج في التنبؤ والإعتماد عليها وإهمال القيم الزائدة والتي يمكن أن تزيد من تشويش النموذج؛

_ بناء النموذج Model Building: يقصد به إختيار الخوارزمية المناسبة وتدريبها من خلال تقسيمها إلى بيانات تدريب Training Data وبيانات إختبار Testing Data وإختيارالنموذج الذي يعطي أكثر دقة في التنبؤ؛

_تقييم النموذج Model Evaluation: بعد بناء النماذج تأتي مرحلة التقييم حيث يتم فيها التحقق من مدى دقة النموذج في التنبؤ والتأكد من أن البيانات التي تدرب عليها النموذج سابقا سيستعملها فعلا للتنبؤ ولن يحتفظ بها فقط، وذلك بالإعتماد على عدة مؤشرات عامة مثل Confusion Matrix وF1 Scour وAccuracy

ثانيا: أدوات تقييم النموذج

بعد تجهيز البيانات لبناء النموذج الخاص بالدراسة يجب أن نتعرف على بعض الأدوات التي سيتم من خلالها تحديد النموذج الأنسب والتي تعتبر كمؤشرات لإختيار أدق النماذج وهي:

أولا : مصفوفة الإلتباس Confusion Matrix: هي عبارة على مصفوفة تقوم بوصف النموذج وتقييمه بطريقة منظمة من خلال مصفوفة الإلتباس يمكن أن نعرف نسبة دقة النموذج وأيضا إكتشاف الأخطاء التي وقع فيها النموذج في مرحلة البناء، وتستعمل عادة في نماذج التصنيف الثنائية ويمكن توضيح مصفوفة التحقق المتقاطع بالشكل التالي:

جدول2.2: *مصفوفة الإرتباك*counfusion matrix

	Actual Positive حقيقة إيجابية	Actual Negative حقيقة سلبية
Predicted Positive توقع إيجابي	True Positive(TP)	False Positive(FP)
Predicted Negative توقع سلبي	False Negative(FN)	True Negative(TN)

المصدر: من إعداد الطالبة بالإعتماد على مخرجات لغة Python

تمثل المصطلحات داخل المصفوفة

- True Positive(TP): تمثل عدد الحالات التي كانت إيجابية والنموذج أيضا قام بتوقعها أنها إيجابية أي الحالات التي تم تصنيفها بشكل صحيح على أنها إيجابية مثل (توقع تخلي الزبائن وقد تخلو فعلا)
- False Negative(FN): تمثل عدد الحالات التي كانت سلبية والنموذج أيضا توقع أنها سلبية أي تم تصنيفها بشكل صحيح على أنها سلبية مثل (توقع بقاء الزبائن وقد بقو فعلا)
- False Positive(FP): تمثل عدد الحالات التي كانت إيجابية لكن النموذج توقع أنها سلبية أي تم تصنيفها بشكل خاطئ على أنها إيجابي مثل (توقع مغادرة الزبائن)
- True Negative(TN): تمثل عدد الحالات التي كانت سلبية لكن النموذج توقع أنها سلبية أي تم تصنيفها بشكل خاطئ على أنها سلبية مثل (توقع بقاء الزبائن)

من خلال هذه العناصر المتعلقة بالمصفوفة يمكن حساب باقي المؤشرات التي تساعد في تقييم النموذج

- ثانيا: معايير إختيار النموذج

لإختيارالنموذج المناسب توجد بعض المعايير والتي يتم حسابها بناءا على مصفوفة التحقق المتقاطع السابقة وتتمثل هذه المعايير في:

1- الدقة Accuracy: يتم إستخدامها لتقييم نماذج التصنيف وهي تعبر عن نسبة التوقعات الصحيحة التي قام بها النموذج ويمكن حسابها من خلال المعادلة التالية:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

حيث:

TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

2- معيار الدقة النوعيةPrecision : تستخدم لتقييم النماذج ثنائية التصنيف وتقيس جودة التوقعات الإيجابية التي يصدرها النموذج ويمكن حسابها كما يلي:

Precision = (True Positives) \ (True Positives+ False Positives)

معيار الإسترجاع Recall: يستخدم لقياس قدرة النموذج على إكتشاف الحالات الحقيقية و يتم حساب معيار الإسترجاع حسب المعادلة التالية: Recall =(TP\TP) + FN

حيث: تمثل TP الحالات الصحيحة التي تنبأ بها النموذج

وتمثل TP + FNجميع الحالات التي اكتشفها النموذج سواء كانت صحيحة أو خاطئة

3- معيار F1 Score: يعتبر هذا المعيار مهما لأنه يجمع بين كل المعيارين السابقة ويهدف إلى تحقيق التوازن بينهم في حالة ما كانت الفئات غير متوازنة التوزيع، يتم حساب هذا المعيار المركب من خلال العلاقة الرياضية التالية: F1 = (2 * (P * R))/ (P + R)

حيث: P تمثل Precision

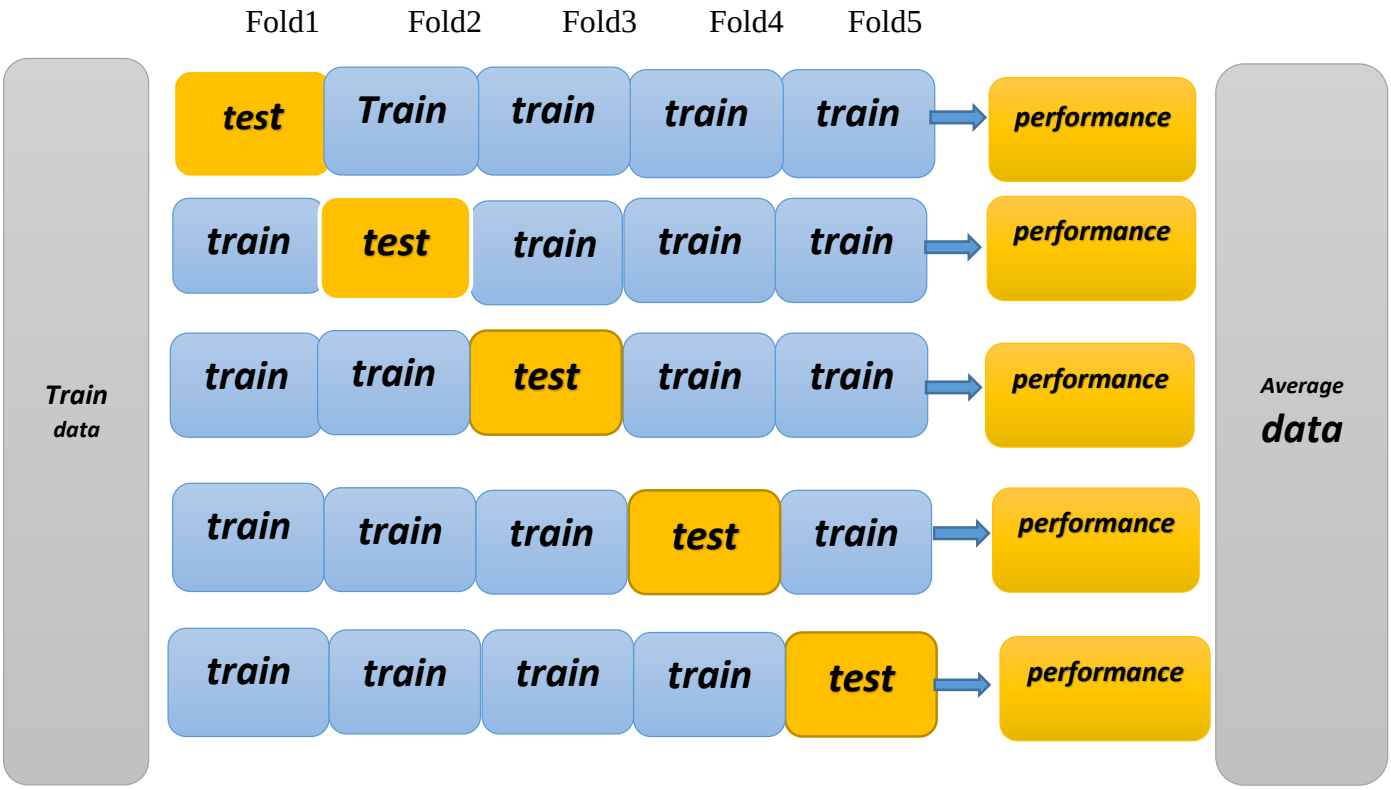
R تمثل Recall

- ثالثا: االتحقق المتقاطع Cross Validation

هي تقنية إحصائية تستخدم في تقييم أداء نماذج التعلم الآلي بهدف ضمان تعميم النتائج على بينات جديدة بحيث تتمثل آلية عمل هذه المصفوفة في أنه بدلا من تقسيم البيانات مرة واحدة إلى بيانات تدريب وبيانات إختبار يتم تقسيم البيانات إلى عدة أجزاء وتسمى بالطيات (Folds) ويتم تدريب النموذج على الطية الأولى وإختباره على الطيات المتبقية والمرحلة الثانية تتغير طية التدريب إلى الطية الثانية وتصبح

الطية الأولى للإختبار وهكذا حتى يتم إختبار كامل الطيات وتدريبها و يمكن ضمان أن كل البيانات تم تدريبها وإختبارها في نفس الوقت أي أن النتائج ستكون أدق أكثر و يمكن شرحها أكثر عبر الرسم التوضيحي التالي:

الشكل9.2: مصفوفة التحقق المتقاطع Cross Validaction



المصدر: من إعداد الطالبة

من خلال الرسم السابق يمكن توضيح مصفوفة الإلتباس بشكل أدق حيث نلاحظ أنه تم تقسيم البيانات على 5 طيات بحيث أنه في أول مرحلة يتم إختبار الطية الأولى (test fold 1) وتدريب باقي بيانات النموذج(Train fold2/ fold3/ fold4/ fold)، أما المرحلة الثانية يتم إختبار الطية الثانية (test fold 2) وتدريب باقي الطيات فتصبح الطية الأولى التي تم إختبارها سابقا تابعة لبيانات التدريب وهكذا حتى يتم إختبار كل الطيات وتدريبها

المبحث الثاني : تحليل النتائج وتفسيرها

بعد التعرف على البيانات و على مصدر جمعها والأدوات التي سنستخدمها في بناء نموذج التنبؤ بسلوك الزبائن المتمثل في بقاء الزبون أو مغادرته وأيضا ذكر أهم المراحل التي سيتم إعتمادها في بناء النموذج تأتي مرحلة تحليل النتائج المتحصل عليها بالإعتماد على برمجية Python وبيئة التطوير

Jupyter Notebook والهدف الأساسي من هذا المبحث هو تفسير النتائج المتحصل عليها من خلال تحليل قاعدة البيانات بإستخدام برمجية Python

المطلب الأول: إستكشاف البيانات و معالجتها

بعد عملية جمع البيانات تأتي مرحلة تحليل البيانات إلا أنه قبل التحليل يجب أن نتأكد من أن البيانات التي سيتم الإعتماد عليها في بناء النموذج نقية وخالية من أي بيانات مفقودة أو متكررة يمكن أن تؤثر على دقة النموذج خاصة وأن البيانات التي سيتم التعامل معها في هذه الدراسة تتكون من 36 متغير و7043 مشاهدة وهذا يعني أنها عرضة لإحتوائها على قيم مفقودة أو بيانات غير متناسقة لهذا يجب تنظيفها

أولا: إستيراد المكتبات الأساسية الخاصة بالبرمجة و التي تتيح لنا إستخذام الأدوات والوظائف المختلفة والفعالة لتنفيذ العمليات المختلفة ويتم إستيراد المكتبات عبر الأوامر التالية:

الشكل10.2: إستيراد المكتبات

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

المصدر: لغة Python

ثانيا : هي إستيراد قاعدة البيانات التي سيتم العمل عليها من خلال الأمر df = pd.read_csv()

الشكل11.2: إستيراد قاعدة البيانات

```
#importation de la base de données brute
df = pd.read_csv('D:/rym/database2025.csv')
```

المصدر: لغة Python

ونظرا لأن قاعدة البيانات كبيرة جدا ولا يمكن حصرها في جدول واحد لذا سأظهر فقط أول 5أسطر من البيانات بإستخدام الأمر df.head()

جدول11.2: عرض أول 5 أسطر من قاعدة البيانات

```
[3]: df.head()
```

[3]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude	Number of Referrals	...	Contract	Paperless Billing	Payment Method	Monthly Charge	Total Charges	Re
0	0002-ORFBO	Female	37	Yes	0	Frazier Park	93225	34.827662	-118.999073	2	...	One Year	Yes	Credit Card	65.6	593.30	
1	0003-MKNFE	Male	46	No	0	Glendale	91206	34.162515	-118.203869	0	...	Month-to-Month	No	Credit Card	-4.0	542.40	
2	0004-TLHLJ	Male	50	No	0	Costa Mesa	92627	33.645672	-117.922613	0	...	Month-to-Month	Yes	Bank Withdrawal	73.9	280.85	
3	0011-IGKFF	Male	78	Yes	0	Martinez	94553	38.014457	-122.115432	1	...	Month-to-Month	Yes	Bank Withdrawal	98.0	1237.85	
4	0013-EXCHZ	Female	75	Yes	0	Camarillo	93010	34.227846	-119.079903	3	...	Month-to-Month	Yes	Credit Card	83.9	267.40	

5 rows × 36 columns

المصدر: المصدر: لغة Python

ثالثا: بعد عرض البيانات المتعلقة بالدراسة يجب معرفة نوع المتغيرات وعددها وهل توجد بها قيم مفقودة أو لا بإستخدام دالة df.info() فتظهر النتائج الموالية:

```
 #   Column                            Non-Null Count Dtype
---  ------                             --------------  -----
 0   Gender                             4835 non-null   int32
 1   Age                                4835 non-null   int64
 2   Married                            4835 non-null   int32
 3   Number of Dependents               4835 non-null   int64
 4   City                               4835 non-null   int32
 5   Zip Code                           4835 non-null   int64
 6   Latitude                           4835 non-null   float64
 7   Longitude                          4835 non-null   float64
 8   Number of Referrals                4835 non-null   int64
 9   Tenure in Months                   4835 non-null   int64
 10 Offer                               4835 non-null   int32
 11 Phone Service                       4835 non-null   int32
 12 Avg Monthly Long Distance Charges   4835 non-null   float64
 13 Multiple Lines                      4835 non-null   int32
 14 Internet Service                    4835 non-null   int32
 15 Internet Type                       4835 non-null   int32
 16 Avg Monthly GB Download             4835 non-null   float64
 17 Online Security                     4835 non-null   int32
 18 Online Backup                       4835 non-null   int32
 19 Device Protection Plan              4835 non-null   int32
 20 Premium Tech Support                4835 non-null   int32
 21 Streaming TV                        4835 non-null   int32
 22 Streaming Movies                    4835 non-null   int32
 23 Streaming Music                     4835 non-null   int32
 24 Unlimited Data                      4835 non-null   int32
 25 Contract                            4835 non-null   int32
 26 Paperless Billing                   4835 non-null   int32
 27 Payment Method                      4835 non-null   int32
 28 Monthly Charge                      4835 non-null   float64
 29 Total Charges                       4835 non-null   float64
 30 Total Refunds                       4835 non-null   float64
 31 Total Extra Data Charges            4835 non-null   int64
 32 Total Long Distance Charges         4835 non-null   float64
 33 Total Revenue                       4835 non-null   float64
dtypes: float64(9), int32(19), int64(6)
```

المصدر: من إعداد الطالبة بناء على مخرجات لغة Python

يتضح من خلال المخرجات السابقة انه لدينا 34 متغير مستقل إضافة إلى المتغير التابع مكونة من 34 متغير مستقل ومتغير تابع يعني إجمالا 35 متغير و توجد متغيرات تحتوي على مشاهدات أقل من 7043 مشاهدة وهذا يدل على وجود قيم مفقودة في المتغيرات التالية:

Internet Type ، Avg Monthly Long Distance Charges ، Avg Monthly Long Distance Charges Devise Protection Plan ،Online Backup ، Online Security ،Avg Monthly GB Download ، Unlimited ، Streaming Music ، Streaming Movies ، Streaming TV،Premium Tech Support، Data

ولتسهيل معرفة البيانات وإستكشاف البيانات بشكل أوضح وأشمل وفهم توزيع القيم وعددها نقوم بعرض المعلومات الإحصائيةحول البيانات والجدول التالي يمثل مجموعة من البانات الإحصائية التي يمكن من خلالها معرفة طبيعة البيانات الاحصائية:

جدول4.2: المعلومات الإحصائية لقاعدة البيانات قبل المعالجة

Number of variables	**36**
Number of observations	7043
Missing cells	20501
Missing cells (%)	8.1%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.9 MiB
Average record size in memory	288.0 B

المصدر: من إعداد الطالبة بالإعتماد على مخرجات لغة Python

نلاحظ من خلال الجدول السابق أنه لدينا:

36 متغير في قاعدة البيانات الكلية قبل إجراء عملية التنظيف والمعالجة ،كما أن عدد المشاهدات بلغ 7043 مشاهدة ، وكان عدد القيم المفقودة 20501 أي بنسبة 8.1% ، إضافة إلى عدم وجود أي صفوف متكررة في قاعدة البيانات بنسبة 0.00%

رابعا : معالجة البيانات

في هذه الخطوة سيتم معالجة البيانات الخام بهدف تحسينها وجعلها اكثر ملائمة للدراسة وبناء نموذج ذو جودة ونسبة دقة مقبولة من خلال معالجة البيانات المفقودة حيث بعد فحص البيانات الكلية وتحديد البيانات المفقودة وبناءا على توزيع هذه القيم في قاعدة البيانات تم إختيار طريقة الحذف كحل لمعالجة البيانات المفقودة وذلك بهدف تجنب أي تأثير سلبي قد يحدث أثناء تدريب النموذج بعد تحديد المتغيرات التي تحتوي على قيم مفقودة وحضر عددها سيتم التعامل معها أيضا بحذف القيم فقط دون حذف المتغيرات ككل وذلك بإستخدام دالة df1= df_clean = df.dropna() ، والجدول التالي يوضح طبيعة البيانات

بعدما تم معالجة مشكلة البيانات المفقودة حيث أصبح عدد المشاهدات 2149 كما بلغ عدد القيم المفقودة 0 بنسبة 0.00% ، إضافة إلى أنه لاتوجد لدينا أي صفوف متكررة

جدول5.2: المعلومات الإحصائية لقاعدة البيانات بعد المعالجة

Number of variables	**36**
Number of observations	4835
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	634.2 KiB
Average record size in memory	296.0 B

المصدر: من إعداد الطالبة بالإعتماد على مخرجات لغة Python

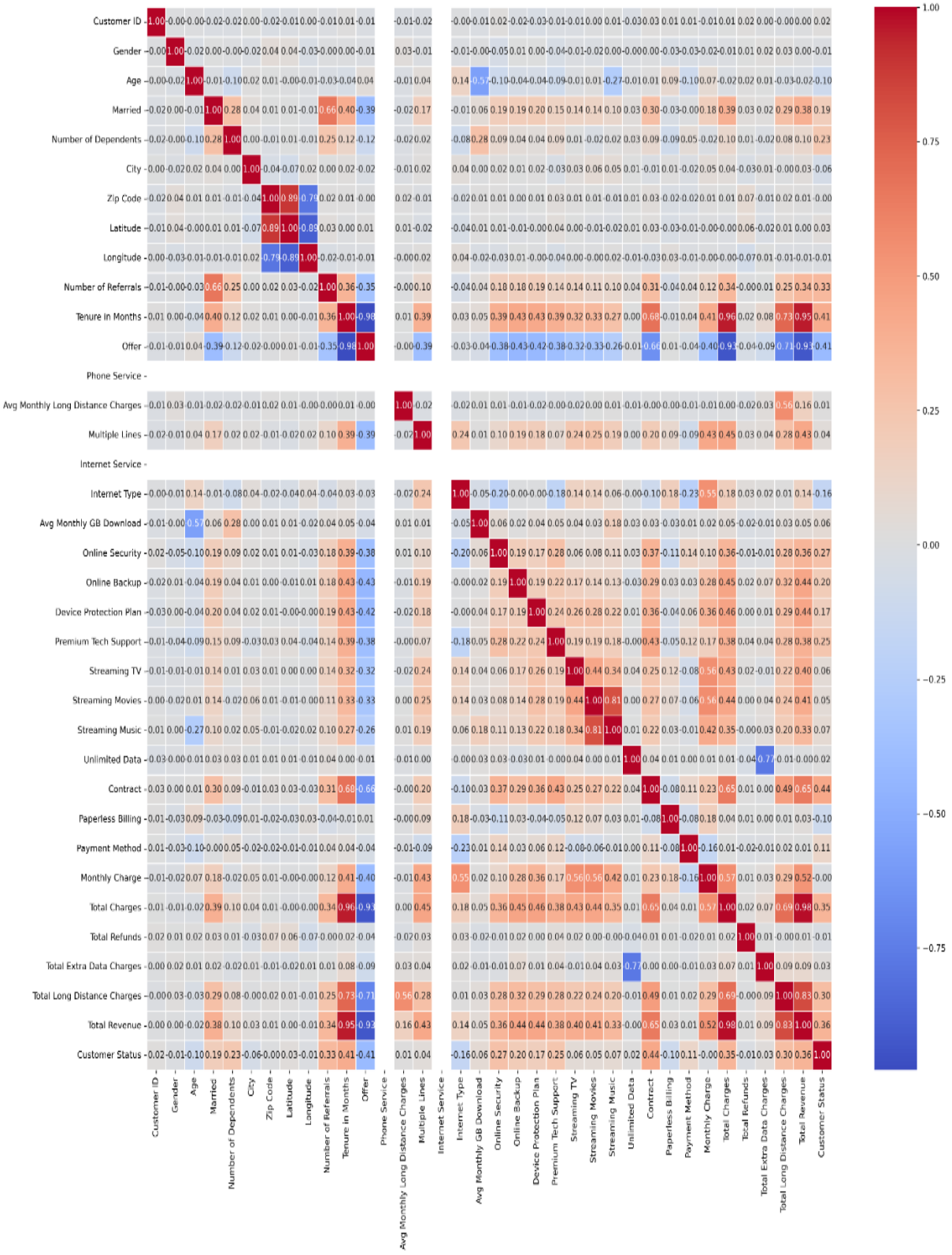
خامسا: ترميز المتغيرات

بما أن خوارزميات التعلم الآلي لا تتعامل إلا مع البيانات الرقمية لا بد من أن تكون جميع البيانات التي سيتم إعتمادها في بناء النموذج رقمية لهذا يجب أن نقوم بتحويل البيانات غير الرقمية والتي بلغ عددها 21 متغير نوعي إلى متغيرات كمية من خلال ترميزها بإستخدام دالة le = LabelEncoder()

سادسا: تحديد المتغيرات

بعد معالجة البيانات وترميزها يجب تحديد المتغيرات المستقلة التي ترتبط إرتباطا قويا بالمتغير التابع فقط و تفادي المتغيرات المستقلة التي لا ترتبط بالمتغير التابع وأيضا المتغيرات المستقلة التي ترتبط ببعضها إرتباطا قويا تجنبا للوقوع في مشكلة الإرتباط الذاتي ولتحديد هذه المتغيرات علينا أن نقوم بمصفوفة الإرتباط الذاتي بإستعمال الأمر والموضحة كالتالي:

الشكل12.2: مصفوفة الإرتباطات



المصدر: لغة Python

من خلال مصفوفة الإرتباطات إتضح أن هناك فقط 28 متغير من بين 35 متغير مستقل لديهم إرتباط قوي بالمتغير التابع وإرتباط ضعيف في ما بينهم ، يمكن شرح الإرتباط حسب الألوان أيضا، حيث يمثل اللون الأحمر الإرتباط التام في قطر المصفوفة والذي يكون بين المتغير ونفسه بمعدل إرتباط يبلغ 1 أما اللون الأزرق في المصفوفة فيمثل الإرتباط القوي والسالب بين المتغيرات ، كما يمثل كل من اللونين الأبيض والرمادي إرتباطا ضعيفا أو معدوم، أما اللون البرتقالي يوضح المتغيرات ذات الإرتباط المتوسط

بعد مصفوفة الإرتباطات علينا أن نقوم بتحديد المتغيرات الأكثرا إرتباطا بالمتغير التابع وأقل إرتباطا في ما بينها من خلال تقنية تسمى wrapper وهي طريقة يتم إستخدامها في عملية features selection

تحديد الميزات أو السمات المهمة والأكثر تأثيرا في بناء نموذج بإستخدام تعلم الآلة حيث تتميز هذه التقنية بدقتها العالية كما أنها تظهر خاصية كل ميزة feature ضمن سياق النموذج أي أنها تأخذ بعين الإعتبار التفاعل بين الخصائص إضافة إلى مرونتهاٌ درتها على التعامل مع كل أنواع النماذج المتغلقة بالتعلم الآلي وتكمن أهمية إستعمال تقنية wrapper في كونها تساعد بشكل كبير على تحسين دقة التنبؤ وذلك بسبب أنها تختلف عن الطرق التقليدية التي تعتمد فقط على العلاقات الإحصائية بين المتغيرات حيث انها تعتمد بالأساس على أداء النموذج الفعلي مما يجعله أكثر واقعية كما أنها تساهم في تقليل الضجيج من البيانات وتتخلص من البيانات غير المؤثرة في النموذج بسرعة عالية والمتغيرات التالية هي التي تم إعتمادها لبناء النموذج والتي تتمثل في 8 متغيرات حيث أن المتغيرtrue تم إختياره ويمكن إعتماده في النموذج أما المتغيرfalse لم يتم إختياره ولا يمكن الإعتماد عليه في بناء النموذج ويمكن معرفتها من خلال النتائج التالية:

الشكل13.2: المتغيرات التي تم إختيارها بالإعتماد على مصفوفة الإرتباطات

```
print('Selected Features are : ' , FeatureSelection.get_support())
```

```
Selected Features are :  [False  True  True  True False False  True False  True  True False False
 False False False False False False False False False False False
 False  True False False  True False False False False False]
```

المصدر: مخرجات لغة Python

ومن خلال إسقاط هذه النتائج على المتغيرات يتبين أن المتغيرات التي تمتلك أكبر علاقة مع المتغير التابع وأضعف علاقة في ما بينها وهي :

Age
Married
Latitude
Number of Referrals
Contract
Monthly Charge
Number of Dependents
Tenure in Months

المطلب الثاني : بناء و تقييم النموذج

بعد الإنتهاء من تحديد المتغيرات المناسبة تأتي أهم مرحلة تهدف إليها هذه الدراسة وهي بناء النموذج الذي سيتنبأ ببقاء الزبائن وإستمرار إستهلاكهم للخدمات التي تقدمها شركة الإتصالات أو مغادرتهم، بناء على المتغيرات التي تم تحديدها مسبقا سنقوم بتدريب عدة نماذج للتنبؤ والمقارنة بينها وإختيار أفضل نموذج حيث تم إستعمال هذه النماذج لتصنيف السلوك الإستهلاكي للزبون نظرا لقدرتهم على التعامل مع البيانات التصنيفية والرقمية وسيتم إعداد هذه النماذج بإسخدام مكتبة Scikit_Learn من لغة Python.

أولا : شجرة القرار Decision Tree

تم إستخدام نموذج شجرة القرار حيث يعتمد هذا النموذج على تقسيم البيانات تدريجيا وفقا للمتغيرات الأكثرا تأثيرا عن طريق حساب معامل الأنتروبي لإختيار أفضل متغير عند كل تقسيم يمكن بناء نموذج شجرة القرار بإستخدام الأمر التالي:

الشكل14.2: إستيراد نموذج شجرة القرار

```
from sklearn.tree import DecisionTreeClassifier
modelDT = DecisionTreeClassifier(random_state=42)
```

المصدر: مخرجات لغة Python

أ- تقييم أداء النموذج بإستخدام مصفوفة التحقق المتقاطعconfusion matrix:

تستخدم مصفوفة االتحقق المتقاطع لتقييم أداء النموذج للتنبؤ بالسلوك الإستهلاكي للزبائن وبما أن الدراسة خاصة بالتصنيف يتم تقسيم السلوك الزبائن إلى فئتين الفئة الأولى هي الفئة التي تقوم بالشراء والفئة الثانية

لا تقوم بالشراء ويمكن إستخراج مصفوفةconfusion matrix الخاصة بنموذج شجرة القرار بإستخدام الأمر التالي:

الشكل 15.2: الأمر الخاص بإستخراج مصفوفة confusion matrix

```
CMDT = confusion_matrix(y_test, y_pred)
```

المصدر: مخرجات لغة Python

حيث كانت نتائج مصفوفة الإلتباس في نموذج شجرة القرار كالتالي:

جدول6.2: نتائج مصفوفة الإلتباس حسب نموذج شجرة القرار

	Actual Positive الزبون غادر فعلا	Actual Negative الزبون بقى فعلا
Predicted Positive توقع الزبون سيغادر	593(TP)	62(FP)
Predicted Negative توقع الزبون سيبقى	101 (FN)	211 (TN)

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

من خلال مصفوفة الإلتباس أن نموذج شجرة القرار توقع أن593 (True Positive) زبون سيشتري وقد إشتروا فعلا وهذا يدل على مدى قدرة النموذج على فهمه الجيد لسلوك الزبائن وميولاتهم وأخطأ في توقع 62 زبون(False Positive) ،كما نلاحظ أن قيمة (False Negative) كانت 101 أي أنه هناك101 زبون توقعهم النموذج بأنهم لم يشترو لكنهم قامو بعملية الشراء و211 زبون كان توقع النموذج بالنسبة لهم صحيح حيث توقع أنهم لن يقوموا بعملية الشراء وفعليا لم يقوموا بذاك(True Negative).

ب- مؤشرات الأداء العامة:

جدول7.2: المؤشرات العامة حسب نموذج شجرة القرار

Accuracy DT	0.83
Precision DT	0.77
Recall DT	0.68
F1 Score DT	0.72

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

بعد بناء نموذج شجرة القرار على بيانات الزبائن تم تقييم أدائه بإستخدام المؤشرات السابقة يتبين أن دقة النموذج العامة Accuracy بلغت نسبة 83% وهي نسبة قوية تدل على الأداء الجيد جدا للنموذج بصفة عامة أي أن النموذج تمكن من تصنيف مانسبته 83% من البيانات بشكل صحيح، كما قدرت الدقة الإيجابية للنموذجPrecision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات الإيجابية الخاطئة بنسبة 77 % وهي نسبة جيدة في التصنيف العام، و بلغت نسبة مؤشر الإسترجاعRecall 68% وهي نسبة مقبولة وتعني أن النموذج توقع نسبة 68% من القيم الإيجابية بشكل صحيح وأخطأ في توقع نسبة 32% من القيم حيث توقعها على أنها قيم إيجابية إلا أنها كانت قيم سالبة، إضافة إلى أن نسبة مؤشر F1 Score بلغت 72% وهي نسبة جيدة تدل على أن النموذج يحقق توازنا بين مؤشر الدقة العامة Accuracy ومؤشر الإسترجاع Recall.

ت- تقييم أداء النموذج بإستخدام مصفوفة التحقق المتقاطع Crosse Validation :

من خلال مصفوفة التحقق المتقاطع الخاصة بنموذج شجرة القرار يمكن أن نقيم أداء نموذج التنبؤ بالسلوك الإستهلاكي للزبون إن كان سيغادر أو سيبقى حيث أضهرت المصفوفة مدى قدرة النموذج على التمييز بين سلوك الزبون المتمثل إما في المغادرة أو البقاء وكما يظهر في المصفوفة التالية أنه تم تقسيم بيانات النموذج على 5 طيات Folds(K=5) وذلك بهدف تدريب البيانات وضمان إستقرار النتائج وتقليل الإنحياز حيث النتائج التالية :

جدول8.2: نتائج مصفوفة التحقق المتقاطع حسب نموذج شجرة القرار

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.7241	0.75801	0.7252	0.72303
Fold2	0.7001	0.7331	0.69843	0.7021
Fold3	0.7035	0.7311	0.6994	0.7103
Fold4	0.74007	0.7714	0.7406	0.7394
Fold5	0.7324	0.76	0.7293	0.7362
Mean cv	0.72	0.75	0.72	0.72

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

يتبين من الجدول السابق الذي يمثل نتائج مصفوفة التحقق المتقاطع والتي تظهر نتائج المؤشرات الأساسية لتقييم نموذج التنبؤ بمغادرة الزبون أو بقائه حسب نموذج شجرة القرار ي أن أداء النموذج جيد ومستقر نسبيا حيث بلغ المتوسط الحسابي لكل من المعايير F1-Scores وPrecision و Recall قيمة 0.72 ومعيار Accuracy قيمة 0,75 %وهذا يشير إلى أن نموذج شجرة القرار يحقق أداءا مقبولا في تصنيف الزبائن إذ أنه يمكن تعميم هذه النتائج بنسبة 72 % إلى % 75، كما أن قيمة الإنحراف المعياري لجميع المؤشرات قدرت ب0,02 وهو إنحراف منخفض و يعكس مدى تقارب الطيات في النموذج وهذا يدل على إستقرار النموذج عبر مختلف الطيات الخمس و يوضح إنخفاض التشتت عبر القيم

ت_ نسبة مساهمة المتغيرات في التنبؤ بالنموذج:

بعد تطبيق نموذج شجرة القرار للتنبؤ بسلوك الزبائن تجاه الشركة إما بالإستمرار في تعاملهم معها أو مغادرتهم وتقييم النموذج بإستخدام كل من مصفوفة التحقق المتقاطع و المؤشرات العامة تم إستخراج نسبة مساهمة كل متغير من المتغيرات المستقلة والتي تم إختيارها سابقا بناءا على إرتباطها القوي مع المتغير التابع وإرتباطها الضعيف في ما بينها في بناء النموذج التنبؤي، حيث تهدف هذه النتائج إلى تحديد العوامل الأكثر تأثيرا على سلوك الزبائن وقرارا الإستمرار أو المغادرة حيث أن هذه الحطوة تساعدج بشكل كبير على معرفة أي من المتغيرات التي يجب التركيز عليها والعمل على تحسينها أكثر بهدف الإحتفاظ بالزبائن.

ث–نسبة مساهمة المتغيرات في بناء نموذج شجرة القرار:

جدول9.2: نسبة مساهمة المتغيرات في بناء نموذج شجرة القرار

المتغير	نسبة المساهمة
Age	%17.22
Married	%00.57
Latitude	%17,14
Number of Referrals	%09.98
Contract	%21,51
Monthly Charge	%17.04
Number of Dependents	%03.82
Tenure in Months	%12,71

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

نلاحظ من الجدول أعلاه أن أكبر نسبة مساهمة كانت في متغيرContract حيث قدرت نسبة مساهمته ب 21,51% مما يوضح أن الزبائن التي تملك عقودا مع الشركة وخاصة تلك التي تكون طويلة الأجل و تصعب مغادرتهم الشركة ويجعلها تحتفظ بتلك النسبة من الزبائن لأكبر وقت ممكن مما يحقق لها دخلا ثابتا، تليها نسب مساهمة المؤشرات Age، Latitude، Monthly Charge وهي على التوالي 17.22%،17,14%، 17.04% وهي نسب متقاربة توضح الدور التكافؤي الذي تلعبه هذه المتغيرات في التأثير على قرارالزبون وهذا يدل على أنه يمكن بناء إستراتيجية شاملة لهذه المتغيرات للحفاظ على الزبائن بدلا من التعامل مع كل متغير لوحده ، كما سجلت نسبة Tenure in Months ب 12,71% وتدل هذه النسبة لتأثير المدة الزمية لإشتراك الزبون في الشركة أي أنه كلما كان إشتراكهم أطول كل ما كان بقاءهم مع الشركة أطول والعكس صحيح، أما نسبة مساهمة المتغير Number of Referrals قدرت ب 09.98% وهي نسبة ملحوضة تشير إلى علاقة البعد الإجتماعي بالزبون وكيفية سلوكه حيث أنه كلما إرتفعت هذه النسبة كلما كان الزبون أكثر ثقة تجاه الشركة وقلت إحتمالية مغادرته، وكانت أقل

نسبة مساهمة لدى المتغيرين Number of Dependents ب 03.82% والمتغير Married بنسبة 00.57% وهي نسب جد ضعيفة تدل على أن حالة الزبائن الإجتماعية لا تؤثر بشكل كبير على سلوكهم تجاه الشركة إلا أنه يمكن أن تؤثر بشكل غير مباشر في النموذج.

ثانيا : الإنحدار اللوجستي

وهو أسلوب إحصائي يمكن إستعماله لتصنيف الزبائن إلى فئتين يعتمد أساسا على حساب إحتمال وقوع أحد فئتي المتغير التابع والمتمثلة في بقاء الزبون أو مغادرته، ويمكن إستيراده في برمجية Python حسب الأمر التالي :

الشكل 16.2: إستيراد نموذج الإنحدار اللوجستي

```
LogisticRegressionModel = LogisticRegression(penalty='l2',solver='sag',C=1.0,random_state=33)
```

المصدر: لغة Python

أ- تقييم أداء النموذج بإستخدام مصفوفة الإلتباس:

يمكن إستخراج مصفوفة الإلتباس الخاصة بنموذج الإنحدار اللوجستي بإستعمال الأمر التالي:

الشكل17.2: أمر مصفوفة الإلتباس الخاصة بنموذج الإنحدار اللوجستي

```
CMLR = confusion_matrix(y_test, y_pred)
CMLR
```

المصدر: لغة Python

والتي كانت نتائجها كالتالي:

جدول10.2: نتائج مصفوفة الإلتباس لنموذج الإنحدار اللوجستي

	Actual Positive الزبون غادر فعلا	Actual Negative الزبون بقي فعلا
Predicted Positive توقع الزبون سيغادر	562 (TP)	93(FP)
Predicted Negative توقع الزبون سيبقى	99 (FN)	213 (TN)

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

من خلال مصفوفة الإلتباس أن نموذج الإنحدار اللوجستي توقع أن 562 (True Positive) زبون سيشتري وقد غادروا فعلا وهذا يدل على مدى قدرة النموذج على فهمه الجيد لسلوك الزبائن وميولاتهم وأخطأ في توقع93 زبون(False Positive) ،كما نلاحظ أن قيمة كانت (False Negative) 99أي أنه هناك 99 زبون توقعهم النموذج بأنهم لم يغادروا لكنهم قامو بالمغادرة و 213 زبون كان توقع النموذج بالنسبة لهم صحيح حيث توقع أنهم لن يقوموا بالمغادرة وفعليا لم يقوموا بذاك(True Negative).

ب- مؤشرات الأداء العامة:

جدول11.2: المؤشرات العامة لنموذج الإنحدار اللوجستي

Accuracy LR	0.80
Precision LR	0.80
Recall LR	0.80
F1 Score LR	0.80

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

بعد تدريب نموذج الإنحدار اللوجستي على بيانات الزبائن تم تقييم أدائه بإستخدام المؤشرات السابقة يتبين أن دقة النموذج العامة Accuracy بلغت نسبة 80% وهي نسبة قوية تدل على الأداء الجيد جدا للنموذج بصفة عامة أي أن النموذج تمكن من توقع السلوك الإستهلاكي للزبائن بنسبة 80% بشكل صحيح، كما قدرت الدقة الإيجابية للنموذجPrecision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات الإيجابية الخاطئة بنسبة80% وهي نسبة جيدة في التصنيف العام، و بلغت نسبة مؤشر الإسترجاعRecall 80%وهي نسبة مقبولة وتعكس قدرة أن النموذج على عدم إهمال أو تفويت الزبائن الحقيقين، إضافة إلى أن نسبة مؤشر F1 Score بلغت 80% وهي نسبة جيدة تدل على أن النموذج يحقق توازنا بين مؤشر الدقة العامة Accuracy ومؤشر الإسترجاع Recall بما أن كل مؤشرات النموذج كانت متساوية فهذا يدل على توازن النموذج وعدم إنحيازه لأي فئة.

ت- تقييم أداء النموذج بإستخدام مصفوفة التحقق المتقاطع Crosse Validation:

جدول12.2: نتائج مصفوفة التحقق المتقاطع حسب نموذج الإنحدار اللوجستي

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.7765	0.8004	0.7733	0.7804
Fold2	0.7369	0.7662	0.7351	0.7388
Fold3	0.7736	0.7962	0.7689	0.7797
Fold4	0.7663	0.7942	0.7665	0.7661
Fold5	0.8055	0.8283	0.8055	0.8055
Mean cv	0.77	0.80	0.77	0.77

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

يتبين من الجدول السابق الذي يمثل نتائج مصفوفة التحقق المتقاطع والتي تظهر نتائج المؤشرات الأساسية لتقييم نموذج التنبؤ بمغادرة الزبون أو بقائه حسب نموذج الإنحدار اللوجستي أن أداء النموذج يتميز بدقة عالية و أكثرا إستقرارا مقارنة بنموذج شجرة القرار وحيث بلغ المتوسط الحسابي لكل من المعاييير F1-Scores وPrecision و Recall قيمة 0.77 ومعيار Accuracy قيمة 0,80 %وهذا يشير إلى أن نموذج الإنحدار اللوجستي يحقق أداءا قويا في تصنيف الزبائن إذ أنه يمكن تعميم هذه النتائج بنسبة 77 % إلى % 80، كما أن قيمة الإنحراف المعياري لجميع المؤشرات قدرت ب0,02 وهو إنحراف منخفض و يعكس مدى تقارب الطيات في النموذج وهذا يدل على إستقرار النموذج عبر مختلف الطيات الخمس و يوضح إنخفاض التشتت عبر القيم.

ث_ نسبة مساهمة كل متغير في النموذج حسب خوارزمية الإنحدار اللوجستي:

جدول13.2: نسبة مساهمة المتغيرات في بناء نموذج الإنحدار اللوجستي

المتغير	نسبة المساهمة
Age	%04.30
Married	%13.95
Latitude	%1.17
Number of Referrals	%26.36
Contract	%19.35
Monthly Charge	%09,60
Number of Dependents	%12.32
Tenure in Months	%12,90

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

يتضح من الجدول أعلاه أن أكبر نسبة مساهمة كانت في متغيرContract حيث قدرت نسبة مساهمته ب 19.35% ، أما المتغير الأقل مساهمة هو متغير Latitude مساهمة بلغت 1.17%

ثالثا: التدرج المعزز Gradient Boosting

هو أحد تقنيات التعلم الآلي يعتمد على بناء عدة نماذج بسيطة مثل أشجار القرار بطريقة تدريجية حيث أن كل نموذج يحسن من الأخطاء التي إرتكبها النموذج الذي قبله ويقوم بتصنيف الزبائن بناءا على الفئة التي تملك أكبر إحتمال ويتم إستيراده من برمجية Python بإستخدام الأمر التالي:

الشكل18.2: أمر إستيراد نموذج التدرج المعزز

```
from sklearn.ensemble import GradientBoostingClassifier
```

المصدر: لغة Python

أ- تقييم أداء النموذج بإستخدام الإلتباس:

يتم إستخراج مصفوفة confusion matrix لنموذج بإستخدام الأمر التالي:

الشكل19.2: إمر إستخراج مصفوفة الإلتباس لنموذج التدرج المعزز

```
CMGBM = confusion_matrix(y_test, y_pred)
print("CM GBM:", CMGBM)
```

المصدر: لغة Python

حيث كانت نتائجها كما يلي:

جدول14.2: نتائج مصفوفة الإلتباس حسب نموذج التدرج المعزز

	Actual Positive الزبون غادر فعلا	Actual Negative الزبون بقي فعلا
Predicted Positive توقع الزبون سيغادر	593 (TP)	62 (FP)
Predicted Negative توقع الزبون سيبقى	101 (FN)	211 (TN)

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

من خلال مصفوفة الإلتباس أن التعزيز التدرجي توقع أن593 (True Positive) زبون سيغادر وقد غادروا فعلا وهذا يدل على مدى قدرة النموذج على فهمه الجيد لسلوك الزبائن وميولاتهم وأخطأ في توقع 62 زبون(False Positive)، كما نلاحظ أن قيمة كانت (False Negative) 101أي أنه هناك 101 زبون توقعهم النموذج بأنهم لن يغادرو لكنهم قامو غادروا 211 زبون كان توقع النموذج بالنسبة لهم صحيح حيث توقع أنهم لن يغادروا وفعليا لم يقوموا بذاك(True Negative)

ب- تفسيير المؤشرات العامة للنموذج:

جدول15.2: المؤاشرات العامة لنموذج التدرج المعزز

Accuracy GBM	0.83
Precision GBM	0.72
Recall GBM	0.68
F1 Score GBM	0.77

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

بعد تدريب نموذج التدرج المعزز على بيانات الزبائن و تم تقييم أدائه بإستخدام المؤشرات السابقة يتبين أن دقة النموذج العامة Accuracy بلغت نسبة 83% وهي نسبةجيدة جدا و تدل على أن النموذج تمكن من توقع السلوك الإستهلاكي للزبائن بنسبة83 % بشكل صحيح ويعكس قدرة النموذج على تقديم نتائج عامة ودقيقة عند تدريبه على بيانات جديدة ، كما سجلت الدقة الإيجابية للنموذج Precision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات الإيجابية الخاطئة بنسبة72 % من التنبؤات التي صنفها النموذج على أنها إيجابية أي أن نموذج التعزيز التدرجي قادر على تقليل التوقعات الخاظئة بنسبة 72% ، و بلغت نسبة مؤشر الإسترجاعRecall 69% وهي نسبة مقبولة وتعكس قدرة أن النموذج على التعرف على السلوك الإستهلاكي للزبائن بنسبة 68% ، إضافة إلى أن نسبة مؤشر F1 Score بلغت 77% وهي نسبة جيدة تدل على أن النموذج يحقق توازنا بين مؤشر الدقة العامة Accuracy ومؤشر الإسترجاع Recall.

ت- تقييم أداء النموذج بإستخدام مصفوفة االتحقق المتقاطع Crosse Validation:

جدول16.2: مصفوفة التحقق المتقاطع لنموذج التدرج المعزز

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.79214663	0.8179938	[0.79416886	0.79027421
Fold2	0.76498011	0.79524302	0.76836554	0.76203834
Fold3	0.80227927	0.82730093	0.80522678	0.79962145
Fold4	0.78285214	0.81178904	0.78826323	0.77838631
Fold5	0.79471228	0.82316443	0.80341632	0.78806776
Mean cv	0.79	0.82	0.79	0.77

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

تظهر نتائج مصقوقة التحقق المتقاطع حسب نموذج التعزيز المتدرج الأداء القوي للنموذج في التنبؤ بسلوك الزبون حيث تراوحت قيمةالمتوسط الحسابي لمؤشرات الأداء بين 0.77 لمعيار Recall وبلغت قيمته بالنسبة لمعياري Precision و F1-Scores 0.79 أما قيمة المتوسط الحسابي بالنسبة لمعيار الدقة العامة Accuracy فقد قدرت ب0.82 وهذا يعكس مدى فعالية النموذج ونجاعته في تصنيف الزبائن حسب مختلف الطيات، كما قدرت قيمة الإنحراف المعياري لنموذج التعزيز المتدرج 0.01 وهو ما يشير إلى مدى تقارب النتائج بين جميع طيات المصفوفة.

ت_ نسبة مساهمة المتغيرات في التنبؤ حسب خوارزمية التدرج االمعزز:

جدول17.2 : نسبة مساهمة المتغيرات في بناء نموذج التدرج المعزز

المتغير	نسبة المساهمة
Age	%08.42
Married	%00.48
Latitude	% 05.64
Number of Referrals	% 07.11
Contract	39.45%
Monthly Charge	% 10.65
Number of Dependents	% 07.10
Tenure in Months	% 11.94

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

تبين من الجدول السابق أن أكبر نسبة مساهمة كانت في متغيرContract حيث قدرت نسبة مساهمته ب 39.45%مما يشير إلى التأثيرالكبير لمتغير العقدContract على تحديد حالة الزبون وسلوكه تجاه الشركة، كما سجلت نسبة مساهمة المتغير Tenure in Months 11.94 % وهي نسبة مرتفعة نسبيا حيث يوضح متغير مدة تعامل الزبون مع الشركة تليها المتغيرات Monthly Chargeو Ageو Numbr of Dependents بنسب مساهمة متقاربة أما المتغيرات Latitudeو Married فقد كانت نسب مساهمتها الأضعف في النموذج.

رابعا: نموذج الغابات العشوائية

تم إعداد نموذج الغابة العشوائية بإستخدام مكتبة Scikt_learn بلغة Python بهدف التنبؤ بالسلوك الإستهلاكي للزبائن في شركة الإتصالات وذلك بناء على المتغيرات المستقلة التي تم التطرق إليها سابقا، حيث إعتمد هذا النموذج على 100 شجرة لبناء نموذج التعلم الآلي بهدف تحسين دقة الأداء والتنبؤ للنموذج من خلال تقوية عدة نماذج مع الإعتماد على ثبات قيمة عشوائية =4 وذلك بهدف إعطاء نفس لنتائج في كل مرة ةتحقيق إستقرار في الأداء عند كل مرة في التجريب بإستخدام الأمر التالي :

الشكل20.2: إستيراد نموذج الغابات العشوائية

```
from sklearn.ensemble import RandomForestClassifier
RFC = RandomForestClassifier(n_estimators=100, random_state=42)
```

المصدر: لغة Python

أ- تقييم الأداء بإستخدام مصفوفة الإلتباس:

يتم إستخراج مصفوفة confusion matrix لنموذج الغابات العشوائية بإستخدام الأمر التالي:

الشكل18.2: أمر إستخراج مصفوفة الإلتباس لنموذج الغابات العشوائية

```
CMRF = confusion_matrix(y_test, y_pred)
CMRF
```

المصدر: لغة Python

وقد كانت نتائج المصفوفة كما يلي:

جدول18.2: نتائج مصفوفة الإلتباس حسب نموذج الغابات العشوائية

	Actual Positive الزبون غادر فعلا	Actual Negative الزبون بقي فعلا
Predicted Positive توقع الزبون سيغاد	591 (TP)	64 (FP)
Predicted Negative توقع الزبون سيبقى	112 (FN)	200 (TN)

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

ب- تفسير المؤشرات العامة للنموذج:

جدول19.2: المؤاشرات العامة لنموذج الغابات العشوائية

Accuracy RF	0.82
Precision RF	0.76
Recall RF	0.64
F1 Score RF	0.69

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

بعد تدريب نموذج االغابات العشوائية على بيانات الزبائن تم تقييم أدائه بإستخدام المؤشرات السابقة حيث تم تقسيم البيانات إلى 5 طيات و تبين أن دقة النموذج العامة Accuracy بلغت نسبة 82% وهي نسبةجيدة جدا و تدل على أن النموذج تمكن من توقع السلوك الإستهلاكي للزبائن ويعكس قدرة النموذج على تقديم نتائج عامة ودقيقة عند تدريبه على بيانات جديدة مما يدل على فعالية نموذج الغابات الغابات العشوائية بشكل عام ، كما سجلت الدقة الإيجابية للنموذج Precision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات الإيجابية الخاطئة بنسبة76% من التنبؤات التي صنفها النموذج على أنها إيجابية أي أن نموذج الغابات العشوائية قادر على تقليل التوقعات الخاطئة بنسبة 76% ، و بلغت نسبة مؤشر الإسترجاع Recall 64% وهي نسبة مقبولة وتعكس قدرة أن النموذج على التعرف على السلوك الإستهلاكي للزبائن بنسبة 64% أي أن أدائه يعتبر مقبول رغم إهماله للنسبة الباقية ، إضافة إلى أن نسبة مؤشر F1 Score بلغت 69% وهي نسبة جيدة تعكس تحقيق التموذج توازنا بين مؤشر الدقة العامة Accuracy ومؤشر الإسترجاع Recall مما يعكس عدم إنحياز النموذج لأي مؤشر من المؤشرين وحسب هذه المؤشرات يعتبر نموذج الغابات العشوائية نموذجا موثوقا ويمكن الإعتماد عليه في التنبؤ بسلوك الزبائن.

ت- تقييم أداء النموذج بإستخدام مصفوفة التحقق المتقاطع Crosse Validation:

جدول20.2: مصفوفة التحقق المتقاطع لنموذج الغابات العشوائية

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.7833	0.8138	0.7921	0.7766
Fold2	0.7595	0.79214	0.7654	0.7541
Fold3	0.7798	0.8086	0.7842	0.7760
Fold4	0.7888	0.8210	0.8039	0.7788
Fold5	0.7755	0.8066	0.7836	0.7693
Mean cv	0.77	0.81	0.78	0.79

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

نلاحظ من خلال النتائج الموضحة في الجدول أعلاه أن الوسيط الخاص بنتائج مؤشرات تقييم نموذج الغابات العشوائية كان 0.77 بالنسبة لمؤشر F1-Scores وبلغ قيمة 0.78 بالنسبة لمؤشر Precision أما قيمة الوسيط لمؤشر Recall فقد كانت 0.79 أما قيمة الوسيط لمؤشر Accuracy 0.81 وهذا يعكس قدرة النموذج على التنبؤ بتصنيف سلوك الزبائن إما بالبقاء أو المغادرة، يعني أنه يمكن تعميم نتائج نموذج التنبؤ بإستخدام خوارزمية الغابات العشوائية بنسبة تتراوح بين 77 % الى غاية 81 %، قيمة الإنحراف المعياري فقد بلغت قيمة 0.01 و وهو ما يشير إلى مدى تقارب النتائج بين جميع طيات المصفوفة.

ث- نسبة مساهمة المتغيرات في التنبؤ بالنموذج حسب خوارزمية الغابات العشوائية :

جدول21.2 : نسبة مساهمة المتغيرات في بناء نموذج الغابات العشوائية

المتغير	نسبة المساهمة
Age	%15.43
Married	%01.86
Latitude	%16.27
Number of Referrals	%09.37
Contract	%13.89
Monthly Charge	%19.12

Number of Dependents	%04.73
Tenure in Months	%19.32

المصدر: من إعداد الطالبة بناءا على مخرجات لغة Python

المطلب الثالث: المقارنة بين الدراسة الحالية والدراسة السابقة

بهدف تأكيد مصداقية النتائج التي تم الحصول عليها في هذه الدراسة ومعرفة الفرق بينها وبين الدراسة السابقة لSumnana Sharma Poudel, Suresh Pokharel, Mohan Timilsina بإعتبارأننا إعتمدنا على نفس قاعدة البيانات telecom_customer_churn.csv سأقوم بإجراء مقارنة بين النتائج التي تحصلت عليها ونتائج الدراسة السابقة والتي تم نشرها سنة 2024 حيث أن الجدول التالي يعطي ملخصا للدراستين الخالية و التي سيتم إجراء مقارنة معها:

جدول22.2: مقارنة بين نتائج الدراستين

الدراسة الحالية	**F1-Scores**	**Accuracy**	**Precision**	**Recall**
-Decision Tree	0.72±0.02	0.02±4750.	0.72±0.02	**0.75 ± 0.01**
-Logistic Regression	0.77±0.01	0.80±0.01	0.77±0.01	**0.77±0.01**
-Gradient Booting	**0.79±0.01**	**0.82±0.01**	**0.79±0.01**	**0.82±0.01**
-Random Forest	0.78±0.01	0.81±0.01	0.79±0.01	**0.77± 0.01**
الدراسة السابقة	F1-Scores	Accuracy	Precision	**Recall**
Logistic Regression	0.54± 0.05	0.79±0.02	0.64±0.04	**0.47±0.06**
Random Forest	0.53±0.07	0.80±0.02	0.71±0.04	**0.43±0.08**
Gradient Booting	0.60±0.02	0.81±0.02	0.67±0.04	0.55±0.03

المصدر :(من إعداد الطالبة بناءا على مخرجات برمجية python و نتائج دراسة Sumnana Sharma (Poudel, SureshPokharel, MohanTimilsina

من خلال الجدول الموضح أعلاه والذي يقدم تلخيصا لأداء كل النماذج التنبؤية التي تم تبنيها في هذه الدراسة إضافة إلى بعض نماذج الدراسة السابقة يمكن ملاحظة نتائج الدراسة الحالية حيث أن أفضل نموذج تم التوصل إليه في الدراسة السابقة وهو نموذج **Gradient Booting** بلغت دقته نسبة 81% بإنحراف معياري قدر ب0.02، أما بالنسبة للدراسة الحالية فقد تبين أن أفضل نموذج من بين كل النماذج التي تم الإعتماد عليها للتنبؤ بمغادرة الزبائن كان أيضا نموذج**Gradient Booting** حيث بلغة نسبة دقته 82% بإنحراف معياري 0.01 ويدل هذا الفرق على تحسن الأداء العام للنموذج من حيث الأداء وأيضا من حيث إستقرار النموذج وثباته والذي يوضحه إنحفاض الإنحراف المعياري، كما أن هناك تباين ملحوظ بين باقي المؤشرات في نموذج **Gradient Booting** في الدراسة السابقة حيث بلغ الفارق نسبة 9% في معيار F1-Score ونسبة 2% حسب معيار Precision أما معيار Recall فقد بلغ الفرق بين الدراستين نسبة27%، وبالرغم من أنه تم إستعمال نفس قاعدة البيانات ونفس النموذج إلا أن هذه الراسة قد سجلت دقة أعلى في التنبؤ بمغادرة الزبائن وإنحرافا معياريا أقل وهذا يشير إلى تحسن في جودة التنبؤ وإستقرار النموذج مما يؤكد صدق النتائج وزيادة نسبة تعميمها، كما لا يفوتنا ملاحظة الفرق بين باقي النمذج سواء من حيث

المؤشرات العامة أو من حيث تظهر قيم الإنحراف المعياري المرتفعة نوعا ما عدم إستقرار النموذج وثباته.

خلاصة الفصل

في هذا الفصل تم التطرق إلى الجانب التطبيقي للدراسة وذلك بالإعتماد على بيانات لشركة إتصالات تم إحضارها من موقع kaggle وذلك لبناء نموذج قادر على التنبؤ يهدف إلى تصنيف الزبائن وتوقع سلوكهم إما المغادرة أو البقاء، حيث تمت معالجة البيانات وتحديد المتغيرات الأكثر ملاءمة لبناء النموذج ثم تطبيق مجموعة من خوارزميات التعلم الآلي التي يعتمد عليها في التنبؤ بهدف هذه الدراسة، كما تم تقييم أداء هذه النموذج حسب عدة مؤشرات بغية معرفة أنسب نموذج من حيث الدقة إضافة إلى إجراء مقارنة بين الدراسة الحالية ودراسة سابقة قد إستخدمت نفس قاعدة البيانات ونفس النموذج تم التوصل إلى أن هذه الدراسة قد تفوقت في نتائجها من حيث دقة التنبؤ وأيضا من حيث إستقرار النموذج.

خاتمة

وفي الختام يمكن القول أن التنبؤ بسلوك الزبائن بالمغادرة Churn يعتبرخطوة جد مهمة في إتخاذ قرارات المؤسسات الإقتصادية والحفاظ على زبائنها وذلك من خلال تحليل البيانات التاريخية المتعلقة بالزبائن، وعتبر تقنيات تعلم الآلة وسيلة جد مناسبة للتعامل مع بيانات الزبائن خاصة وأنها عادة ما تكون هذه البيانات متنوعة تشمل بيانات ديموغرافية وإقتصادية،...إلخ وقد سعيت في هذا البحث إلى بناء نموذج تنبؤي قادر على تصنيف الزبائن بدقة بالإعتماد على بياناتهم بهدف معرفة سلوكهم مبكرا إذ تبين من خلال تجريب أكثر من نموذج تصنيف أن نموذج التدرج المعزز كان أفضل نموذج من حيث دقة التنبؤ وإستقرار البيانات، تؤكد هذه الدراسة على أهمية إستعمال تقنيات تعلم الآلة في تحليل البيانات وكشف الأنماط الخفية في سلوك الزبائن، مما يحث المؤسسات على التوجه نحو إستعمال تقنيات تعلم الآلة بدلا من الطرق التقليدية.

إختبار الفرضيات:

- أثبتت نتائج الدراسة أن أداء النموذجين الغابات العشوائية Random Forest والتدرج المعزز GBM يفوق أداء نموذج شجرة القرار Decision Tree و نموذج الإنحدار اللوجستي Logistic Regression في تصنيف سلوك الزبون الاستهلاكي، حيث حققت هذه النماذج نسب أعلى في مؤشراتF1-Scores وAccuracy و Precisionو Recall وهذا مايثبت صحة الفرضية.

- إستخدام تقنية Cross Validation ساعد على تقليل تأثير تقسيم البيانات العشوائ على النتائج مما يعطي أداء أكثر إستقرارا ودقة، حيث أظهرت تنتائج التقييم أن الأداء كان أقل تذبذبا وأكثر إستقرارا وثباتا وهذا يدل على صحة الفرضية.

- من خلال نتائج الدراسة تبين أن نموذج التدرج المعزز GBM قد حقق أعلى أداء مقارنة بباقي النماذج المستخدمة في مهمة تصنيف الزبائن وفقا لسلوكهم الإستهلاكي، وذلك من حيث المؤشرات العامة وهو ما يؤكد صحة الفرضية.

نتائج الدراسة:

- الزبائن الذين يملكون عقودا طويلة الأجل مع الشركة هم أقل الزبائن مغادرة

- تعتبر نماذج شجرة القرار والإنحدار اللوجستي والتعزز المدرج والغابات العشوائية من بين أهم نماذج التصنيف التي يمكن إستخدامها لتصنيف سلوك الزبائن.

- أثبتت خطوة معالجة البيانات وتنظيفها أنه ليس بالضرورة أن تكون كل المتغيرات في قاعدة البيانات صالحة لبناء النموذج، بل يبنى نموذج التنبؤ على أساس المتغيرات المستقلة الأكثر إرتباطا بالمتغير التابع والأقل إرتباطا في ما بينها مما يؤدي إلى عدم الوقوع في مشكلة الإرتباط الذاتي بين الأخطاء.

-من خلال المقارنة بين نماذج تعلم الآلة المستخدمة تم التوصل إلى أن نموذجGBM التدرج المعززهو أفضل نموذج يمكن إستخدامه لتصنيف الزبائن كما أن التوازن بين معيار الدقة والإسترجاع في نموذج GBM بلغت نسبة 79 % وإنحراف معياري 0.01 وهو ما يؤكد كفاءة النموذج في التصنيف

-أكدت نتائج الدراسة أن إستخدام تقنيات تعلم الآلة وخاصة خوارزميات التصنيف تعتبرأداة ذات كفاءة لبناء نموذج تنبؤ بمغادرة الزبائن خاصة إذا تم دعمها بتقنيات تقييم موثوقة مثل تقنية Cross Validation

توصيات الدراسة:

-تقديم حوافز للزبائن بهدف تشجيعهم عل توقيع العقود طويلة الأجل بغية الإحتفاظ بهم؛

-عند إجراء التنبؤ بإستخدام تقنيات تعلم الآلة يفضل تجريب عدة نماذج بدلا من الإعتماد على نموذج واحد ومقارنتهم وإختيار النموذج الذي يعطي أكثر دقة وأكثر إستقرارا؛

-الإعتماد على نتائج نموذج التنبؤ في إتخاذ قرارات المؤسسة الإقتصادية بهدف معرفة الزبائن المعرضين للمغادرة بشكل مسبق مما يساعد على العمل على التحسين من تجربتهم وكسب ولائهم؛

-إستخدام تقنيات تعلم الآلة للتنبر بالسلوك الإستهلاكي للزبون كونها تملك القدرة على التعامل مع بيانات كبيرة ومتنوعة وتحليلها في وقت قياسي.

قائمة المراجع

arkam. (2017, 17 مارس). *الفرق بين الزبون والعميل والفرق بين المستهلك والمشتري*. Retrieved from arkam: https://arkam.ly

Cutluer, A. c. (2011, January). Random forests. p. 1.

Haden. (2023, 12 29). *Medium*. Retrieved from https://images. app. goo.gl /Xn8zvm1qJ1SJKzxN7

Hesham Moussa. (n.d.). *المقالة*. Retrieved from https://heshammoussa.com/

koleva, D. (2024). Machine learning methods applicable in customer lifecycle management . *Izvestia Journal Of The Union Of Scientists _VARNA*, 1.

raja, V., & Jumes Kung. (2025). "Predicting Customer Behavior in E-Commerce Using Machine Learning Algorithms: A Mathematical Approach. 15678.

Redzebra, A. V. (2016). Predicting online user behaviour using deep learning algorithms.

Seaborn. (2025). *تصور البيانات الإحصائية*. Retrieved from Seaborn: https://seaborn.pydata.org/tutorial/introduction

Sharma, M. K., & Vaibhav Patel, A. (2023). Machin learning based customer churn predaction using improved feature selection techniques. *International Research Journal of Engineering & Applied Sciences (IRJEAS)*, 11.

Shurma, S., & Suresh Pokharel, M. (2024). Explaining Customer Churn Prediction in Telecom Industry Using Tabular Machin Learning Models. *Machin Learning With Aplication*, 17.

Taher, H. M. (n.d). *Learn python for biginners* . oulabook.

Trevor Hastie, & Robert Tibshirani, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* California: Sprenger.

(2025, 01 31). *إستعراف*. Retrieved from الطبي: https://www.bing.com

الجريسي، خ. ب. (1427ه). *سلوك المستهلك* . السعودية.

الطائي، ي. ح., & هشام فوزي العبادي. (2009). *إدارة علاقة الزبائن*. عمان: الوراق للنشر والتوزيع.

العبيد، ع. ا. (2003). *مبتدئ التنبؤ الإداري*. السعودية: النشر العلمي والمطابع.

.*مفهوم العميل وأنواع العملاء والفرق بين العميل والزبون والمستهلك* .(اكتوبر 24 ,2023) .الغامدي، ا
Retrieved from مفاهيم: https://mafaheem.info

.*المفاضلة بين نموذج السلاسل الزمنية ونموذج الإنحدار البسيط* .(2006/2005)

دور تحليلات البيانات الضخمة في التنبؤ بينية تحول .(2024) .حسين مصيلحي سيد، إ & .,النجار، ط
.17 ,*مجلة الدراسات التجارية المعاصرة* .العملاء بإستخدام تحليل المشاعر

,أ. ن. العشعوش) .*التنبؤ بالمبيعات بين النظرية والتطبيق* .(ه1429) .ريجي بولوني، ج. ك & ,إيزينيه
.الرياض (.Trans

سكيكدة: جامعة 20 أوت . *محاظرات في مقياس إدارة العلاقة مع الزبون* .(2020_2019) .إيمان، خ. أ
.1995 سكيكدة

.ورقلة .*محددات سلوك المستهلك* .(2023) .سعاد، ب

.*التسويق للجميع* .(2009) .شبايك، ر

.الكويت: عالم المعرفة .*التنبؤ العلمي ومستقبل الانسان* .(1978) .صالح، ع. ا

مجلة .تأثير إستخدام أدوات الذكاء الإصطناعيفي التنبؤ بسلوك المستهلك .(2024) .عامر، ن. س
.70 ,*البحوث الإعلامية*

.*تعلم الآلة* .(2018) .عثمان، ع. ا

.*دور تكنولوجيا المعلومات والإتصال في تفعيل إدارة علاقة الزبون* .(2010/2011) .كنزة، سكر

.*مدخل الى الذكاء الاصطناعي وتعلم الالة* .(2020) .لحلح، م

المقارنة ما بين تعلم الآلة والتعلم العميق في بناء نموذج قادر على التنبر .(2023) .محمود، ح. خ
.2 ,*المجلة المصرية للسكان وتنظيم الأسرة* .بخروج العملاء من الشركات الخدمية

.سوريا: الجامعة الإفتراضية السورية .*سلوك المستهلك* .(2020) .معراوي، أ

المقارنة ما بين تعلم الآلة والتعلم العميق في بناء نموذج قادر على التنبؤ .(17 8 ,2003) .مهدي، ح. خ
.المجلة المصرية للسكان وتنظيم الأسرة .p .بخروج العملاء من الشركات الخذمية

علوم الإقتصاد والتسيير . تأثير مستوى الرضا لدى الزبون في زيادة ولائه للعلامة .(.n.d) .نجود، ح
.*والتجارة*

. قالمة .*دور الإشهار الإذاعي المحلي في التأثير على سلوك المستهلك* .(2019/2020) .هدى، م. ا

.العراق .*التعلم العميق من الأساسيات ، حتى بناء شبكة عصبية عميقة بلغة البايثون* .(2022) .وازن، م

.العراق .*تعلم الآلة وعلم البانات* .(.n.d) .وازن، م